

12/31/97

Express Mail" mailing label number EM 062 546 604 US. I hereby certify that this document and referenced attachments are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR § 1.10, addressed to: Assistant Commissioner for Patents, Box Sequence, Washington, D.C. 20231 on December 31, 1997.

By: Emma Durrell  
Printed: Emma Durrell

Assistant Commissioner for Patents  
Box Sequence  
Washington, D.C. 20231

Transmitted herewith for filing is the patent application of:

Inventors: Preeti Lal, Jennifer L. Hillman, Neil C. Corley, Karl J. Guegler, Mariah Baugh, Susan Sather and Purvi Shah

Title: **HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS**

Enclosed are:

- ☒ Return postcard,
- ☒ 117 Pages of Specification;
- ☒ 169 Pages of Sequence Listing.
- ☒ 3 Pages of Claims,
- ☒ 1 Page of Abstract;
- ☒ -0- Pages of Figures
- ☒ 5 Pages - **Unexecuted** Declaration and Power of Attorney; and
- ☒ 1 Page of Sequence Listing Statement and one (1) Computer-Readable Diskette.

**Fee Calculation** - The fee has been calculated as follows:

**CLAIMS AS FILED** (fees computed under § 1.9(f))

Claims	Number Filed	Minus	Number Extra	Other Than Small Entity Rate	Basic Fee
					\$790.00
Total Claims	23	-20	3	x \$22	\$ 66 00
Indep. Claims	2	-3	-0-	x \$82	\$ -0-
Multiple Dependent Claim(s), if any + \$270					\$ -0-

**TOTAL FILING FEE \$ 856.00**

The Commissioner is hereby authorized to charge Deposit Account No. **09-0108** in the amount of **\$ 856.00**

The Commissioner is hereby authorized to charge any additional fees required under 37 C.F.R. § 1.16 and 1.17, or credit any overpayment to Deposit Account No. **09-0108**. A duplicate of this sheet is enclosed.

Respectfully submitted,

INCYTE PHARMACEUTICALS, INC.

Date: December 31, 1997

Sheela Mohan-Peterson  
Reg. No. 41,201

3174 Porter Drive  
Palo Alto, California 94304  
Phone: (650) 855-0555  
Fax: (650) 845-4166

"Express Mail" mailing label number EM 062 546 604 US. I hereby certify that this document and referenced attachments are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR § 1.10, addressed to: Assistant Commissioner for Patents, Box Sequence, Washington, D.C. 20231 on December 31, 1997.

By: Emma Durrell

Printed: Emma Durrell

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Lal et al.

Title: **HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS**

Serial No.: To Be Assigned

Filing Date: Herewith

Examiner: To Be Assigned

Group Art Unit: To Be Assigned

Assistant Commissioner for Patents

**Box Sequence**

Washington, D.C. 20231

**SUBMISSION UNDER 37 CFR § 1.821-1.825 SEQUENCE LISTING**

Sir:

In accordance with the requirements of 37 CFR § 1.821-1.825, Applicants hereby submit one diskette containing the computer-readable information for the Sequence Listing of the above-identified application. The diskette complies with the requirements of 37 CFR § 1.824 and is IBM PC compatible using a PC-DOS/MS-DOS 6.2 operating system with Fastseq software version 2.0.

Contained within the application, as filed, just before the claim section, is the Sequence Listing paper copy of the sequences disclosed in the application.

The content of the Sequence Listing paper copy is identical to the computer-readable copy, as required under 37 CFR § 1.821(f).

Respectfully submitted,

INCYTE PHARMACEUTICALS, INC.

Date: December 31, 1997

Sheela Mohan-Peterson  
Sheela Mohan-Peterson  
Reg. No. 41,201

3174 Porter Drive  
Palo Alto, California 94304  
Phone: (650) 855-0555  
Fax: (650) 845-4166

## HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS

### FIELD OF THE INVENTION

This invention relates to nucleic acid and amino acid sequences of human signal peptide-containing proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cancer and immunological disorders.

### BACKGROUND OF THE INVENTION

Protein transport is an essential process for all living cells. Transport of an individual protein usually occurs via an amino-terminal signal sequence which directs, or targets, the protein from its ribosomal assembly site to a particular cellular or extracellular location. Transport may involve any combination of several of the following steps: contact with a chaperone, unfolding, interaction with a receptor and/or a pore complex, addition of energy, and refolding. Moreover, an extracellular protein may be produced as an inactive precursor. Once the precursor has been exported, removal of the signal sequence by a signal peptidase and posttranslational processing (e.g., glycosylation or phosphorylation) activates the protein. Signal sequences are common to receptors, matrix molecules (e.g., adhesion, cadherin, extracellular matrix, integrin, and selectin), cytokines, hormones, growth and differentiation factors, neuropeptides, vasomediators, phosphokinases, phosphatases, phospholipases, phosphodiesterases, G and Ras-related proteins, ion channels, transporters/pumps, proteases, and transcription factors.

G-protein coupled receptors (GPCRs) are a superfamily of integral membrane proteins which transduce extracellular signals. GPCRs include receptors for biogenic amines, e.g., dopamine, epinephrine, histamine, glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin; for lipid mediators of inflammation such as prostaglandins, platelet activating factor, and leukotrienes; for peptide hormones such as calcitonin, C5a anaphylatoxin, follicle stimulating hormone, gonadotropin releasing hormone, neurokinin, oxytocin, and thrombin; and for sensory signal mediators, e.g., retinal photopigments and olfactory stimulatory molecules.

The structure of these highly-conserved receptors consists of seven hydrophobic transmembrane regions, cysteine disulfide bridges between the second and third extracellular loops, an extracellular N-terminus, and a cytoplasmic C-terminus. Three extracellular loops alternate with three intracellular loops to link the seven transmembrane regions. The

5 N-terminus interacts with ligands, the disulfide bridge interacts with agonists and antagonists, and the large third intracellular loop interacts with G proteins to activate second messengers such as cyclic AMP (cAMP), phospholipase C, inositol triphosphate, or ion channel proteins. The most conserved parts of these proteins are the transmembrane regions and the first two cytoplasmic loops. A conserved, acidic-Arg-aromatic triplet present in the second

10 cytoplasmic loop may interact with the G proteins. The consensus pattern, [GSTALIVMYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)-[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)-[LIVM] is characteristic of most proteins belonging to this superfamily. (Watson, S. and Arkininstall, S. (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego, CA, pp. 2-6; and Bolander, F.F. 15 (1994) Molecular Endocrinology, Academic Press, San Diego, CA, pp. 8-19.)

Tetraspanins are a superfamily of membrane proteins which facilitate the formation and stability of cell-surface signaling complexes containing lineage-specific proteins, integrins, and other tetraspanins. They are involved in cell activation, proliferation (including cancer), differentiation, adhesion, and motility. These proteins cross the membrane four 20 times, have conserved intracellular N- and C-termini and an extracellular, non-conserved hydrophilic domain. Three highly conserved polar amino acids are located in the transmembrane domains (TM), an asparagine in TM1 and a glutamate or glutamine in TM3 and TM4. Two to three conserved charged residues, including a glutamic acid residue, are present in the cytoplasmic loop between TM2 and TM3. The extracellular loop between 25 TM3 and TM4 contains four conserved cysteine residues: two in a conserved CCG motif located about 50 residues C-terminal to TM3; one, often preceded by glycine, 11 residues N-terminal to TM4; and one in the extracellular loop may be found in a PXSC motif.

Tetraspanins include, e.g., platelet and endothelial cell membrane proteins, leukocyte surface proteins, tissue specific and tumorous antigens, and the retinitis pigmentosa-associated gene 30 peripherin. (Maecker, H.T. et al. (1997) FASEB J. 11:428-442.) Matrix proteins (Mps)



function in formation, growth, remodeling and maintenance of tissues and as important mediators and regulators of the inflammatory response. The expression and balance of MPs may be perturbed by biochemical changes that result from congenital, epigenetic, or infectious diseases. In addition, MPs affect leukocyte migration, proliferation, differentiation, and activation in immune response.

MPs encompass a variety of proteins and their functions. Extracellular matrix (ECM) proteins are multidomain proteins that play an important role in the diverse functions of the ECM. ECM proteins are frequently characterized by the presence of one or more domains which may include collagen-like domains, EGF-like domains, immunoglobulin-like domains, fibronectin-like domains, vWFA-like modules. (Ayad, S. et al. (1994) The Extracellular Matrix Facts Book, Academic Press, San Diego, CA, pp. 2-16.) Cell adhesion molecules (CAMs) have been shown to stimulate axonal growth through homophilic and/or heterophilic interactions with other molecules. In addition, interactions between adhesion molecules and their receptors can potentiate the effects of growth factors upon cell biochemistry via shared signaling pathways. (Ruoslahti, E. (1997) *Kidney Int.* 51:1413-1417.) Cadherins comprise a family of calcium-dependant glycoproteins that function in mediating cell-cell adhesion in solid tissues of multicellular organisms. Integrins are ubiquitous transmembrane adhesion molecules that link cells to the ECM by interacting with the cytoskeleton. Integrins also function as signal transduction receptors and stimulate changes in intracellular calcium levels and protein kinase activity. (Sjaastad, M.D. and Nelson, W.J. (1997) *BioEssays* 19:47-55.)

Lectins are proteins characterized by their ability to bind carbohydrates on cell membranes by means of discrete, modular carbohydrate recognition domains, CRDs. (Kishore, U. et al. (1997) *Matrix Biol.* 15:583-592.) Certain cytokines and membrane-spanning proteins have CRDs which may enhance interactions with extracellular or intracellular ligands, with proteins in secretory pathways, or with molecules in signal transduction pathways. The lipocalin superfamily constitutes a phylogenetically conserved group of more than forty proteins that function by binding to and transporting a variety of physiologically important ligands. Members of this family function as carriers of retinoids, odorants, chromophores, pheromones, and sterols, and a subset of these proteins may be multifunctional, serving as either a biosynthetic enzyme or as a specific enzyme inhibitor.

(Tanaka, T. et al. (1997) J. Biol. Chem. 272:15789-15795; and van't Hof, W. et al. (1997) J. Biol. Chem. 272:1837-1841.) Selectins are a family of calcium ion-dependent lectins expressed on inflamed vascular endothelium and the surface of some leukocytes. They mediate rolling movement and adhesive contacts between blood cells and blood vessel walls.

5 The structure of the selectins and their ligands supports the type of bond formation and dissociation that allows a cell to roll under conditions of flow. (Rossiter, H. et al. (1997) Mol. Med. Today 3:214-222.)

Protein kinases regulate many different cell proliferation, differentiation, and signaling processes by adding phosphate groups to proteins. Reversible protein

10 phosphorylation is a key strategy for controlling protein functional activity in eukaryotic cells. The high energy phosphate which drives this activation is generally transferred from adenosine triphosphate molecules (ATP) to a particular protein by protein kinases and removed from that protein by protein phosphatases. Phosphorylation occurs in response to extracellular signals, cell cycle checkpoints, and environmental or nutritional stresses.

15 Protein kinases may be roughly divided into two groups; protein tyrosine kinases (PTKs) which phosphorylate tyrosine residues, and serine/threonine kinases (STKs) which phosphorylate serine or threonine residues. A few protein kinases have dual specificity. A majority of kinases contain a similar 250-300 amino acid catalytic domain which can be further divided into eleven subdomains. The N-terminal domain, which contains subdomains

20 I to IV, generally folds into a two-lobed structure which binds and orients the ATP (or GTP) donor molecule. The larger C terminal domain, which contains subdomains VIA to XI, binds the protein substrate and carries out the transfer of the gamma phosphate from ATP to the hydroxyl group of the target amino acid residue. Subdomain V links the two domains. Each of the 11 subdomains contain specific residues and motifs that are characteristic and are

25 highly conserved. (Hardie, G. and Hanks, S. (1995) The Protein Kinase Facts Book, Vol I, pp. 7-47, Academic Press, San Diego, CA.)

Protein phosphatases remove phosphate groups from molecules previously modified by protein kinases thus participating in cell signaling, proliferation, differentiation, contacts, and oncogenesis. Protein phosphorylation is a key strategy used to control protein functional

30 activity in eukaryotic cells. The high energy phosphate is transferred from ATP to a protein

by protein kinases and removed by protein phosphatases. There appear to be three, evolutionarily-distinct protein phosphatase gene families: protein phosphatases (PPs); protein tyrosine phosphatases (PTPs); and acid/alkaline phosphatases (APs). PPs dephosphorylate phosphoserine/threonine residues and are an important regulator of many cAMP mediated, hormone responses in cells. PTPs reverse the effects of protein tyrosine kinases and therefore play a significant role in cell cycle and cell signaling processes. Although APs dephosphorylate substrates in vitro, their role in vivo is not well known. (Carbonneau, H. and Tonks, N.K. (1992) *Annu. Rev. Cell Biol.* 8:463-493.)

Protein phosphatase inhibitors control the activities of specific phosphatases. A specific inhibitor of PP-I, I-1, has been identified that when phosphorylated by cAMP-dependent protein kinase (PKA) specifically binds to PP-I and inhibits its activity. Since PP-I is dephosphorylates many of the proteins phosphorylated by PKA, activation of I-1 by PKA serves to amplify the effects of PKA and the many cAMP-dependent responses mediated by PKA. In addition, since PP-I also dephosphorylates many phosphoproteins that are not phosphorylated by PKA, I-1 activation serves to exert cAMP control over other protein phosphorylations. I<sub>1</sub>PP2A is a specific and potent inhibitor of PP-IIA. (Li, M. et al. (1996) *Biochemistry* 35:6998-7002.) Since PP-IIA is the main phosphatase responsible for reversing the phosphorylations of serine/threonine kinases, I<sub>1</sub>PP2A has broad effects in controlling protein phosphorylations.

Cyclic nucleotides (cAMP and cGMP) function as intracellular second messengers to transduce a variety of extracellular signals, including hormones, and light and neurotransmitters. Cyclic nucleotide phosphodiesterases (PDEs) degrade cyclic nucleotides to their corresponding monophosphates, thereby regulating the intracellular concentrations of cyclic nucleotides and their effects on signal transduction. At least seven families of mammalian PDEs have been identified based on substrate specificity and affinity, sensitivity to cofactors and sensitivity to inhibitory drugs. (Beavo, J.A. (1995) *Physiological Reviews* 75: 725-748.) PDEs are composed of a catalytic domain of ~270 amino acids, an N-terminal regulatory domain responsible for binding cofactors and, in some cases, a C-terminal domain with unknown function. Within the catalytic domain, there is approximately 30% amino acid identity between PDE families and ~85-95% identity between isozymes of the same family.

Furthermore, within a family there is extensive similarity (>60%) outside the catalytic domain, while across families there is little or no sequence similarity. A variety of diseases have been attributed to increased PDE activity and inhibitors of PDEs have been used effectively as anti-inflammatory, antihypertensive, and antithrombotic agents. (Verghese, M.W. et al. (1995) Mol. Pharmacol. 47:1164-1171; and Banner, K.H.. and Page, C.P. (1995) Eur. Respir. J. 8:996-1000.)

Phospholipases (PLs) are enzymes that catalyze the removal of fatty acid residues from phosphoglycerides. PLs play an important role in transmembrane signal transduction and are named according to the specific ester bond in phosphoglycerides that is hydrolyzed, i.e., A<sub>1</sub>, A<sub>2</sub>, C or D. PLA<sub>2</sub> cleaves the ester bond at position 2 of the glycerol moiety of membrane phospholipids giving rise to arachidonic acid. Arachidonic acid is the common precursor to four major classes of eicosanoids; prostaglandins, prostacyclins, thromboxanes and leukotrienes. Eicosanoids are signaling molecules involved in the contraction of smooth muscle, platelet aggregation, and pain and inflammatory responses. PLC is an important link in certain receptor-mediated, signaling transduction pathways. Extracellular signaling molecules including hormones, growth factors, neurotransmitters, and immunoglobulins bind to their respective cell surface receptors and activate PLC. Activated PLC generates second messenger molecules from the hydrolysis of inositol phospholipids that regulate cellular processes, e.g., secretion, neural activity, metabolism and proliferation. (Alberts, B. et al. (1994) Molecular Biology of The Cell, Garland Publishing, Inc., New York, NY, pp. 85, 211, 239-240, 642-645.)

The nucleotide cyclases, i.e., adenylate and guanylate cyclase, catalyze the synthesis of the cyclic nucleotides, cAMP and cGMP, from ATP and GTP, respectively. They act in concert with phosphodiesterases, which degrade cAMP and cGMP, to regulate the cellular levels of these molecules and their functions. cAMP and cGMP function as intracellular second messengers to transduce a variety of extracellular signals, e.g., hormones, and light and neurotransmitters. Adenylate cyclase is a plasma membrane protein that is coupled with various hormone receptors also located on the plasma membrane. Binding of a hormone to its receptor activates adenylate cyclase which, in turn, increases the levels of cAMP in the cytosol. The activation of other molecules by cAMP leads to the cellular effect of the

hormone. In a similar manner, guanylate cyclase participates in the process of visual excitation and phototransduction in the eye. (Stryer, L. (1988) Biochemistry W.H. Freeman and Co., New York, pp. 975-980, 1029-1035.) Cytokines are produced in response to cell perturbation. Some cytokines are produced as precursor forms, and some form multimers in order to become active. They are produced in groups and in patterns characteristic of the particular stimulus or disease, and the members of the group interact with one another and other molecules to produce an overall biological response. Interleukins, neurotrophins, growth factors, interferons, and chemokines are all families of cytokines which work in conjunction with cellular receptors to regulate cell proliferation and differentiation and to affect such activities, e.g., leukocyte migration and function, hematopoietic cell proliferation, temperature regulation, acute response to infections, tissue remodeling, and cell survival. Studies using antibodies or other drugs that modify the activity of a particular cytokine are used to elucidate the roles of individual cytokines in pathology and physiology.

Chemokines are a small chemoattractant cytokines which are active in leukocyte trafficking. Initially, chemokines were isolated and purified from inflamed tissues, but recently several chemokines have been discovered through molecular cloning techniques. Chemokines have been shown to be active in cell activation and migration, angiogenic and angiostatic activities, suppression of hematopoiesis, HIV infectivity, and promoting Th-1(IL-2-, interferon  $\gamma$ -stimulated) cytokine release.

Chemokines generally contain 70-100 amino acids and are subdivided into four subfamilies based on the presence and arrangement of conserved CXC, CC, CX3C and C motifs. The CXC (alpha), CC (beta), and CX3C chemokines contain four conserved cysteines. The CC subfamily is active on monocytes, lymphocytes, eosinophils, and mast cells; the CXC subfamily, on neutrophils; CX3C and C subfamilies, on T-cells. Many of the CC chemokines have been characterized functionally as well as structurally. (Callard, R. and Gearing, A. (1994) The Cytokine Facts Book, Academic Press, New York, NY, pp. 181-190, 210-213, 223-227.)

Growth and differentiation factors function in intercellular communication. Once secreted from the cell, some factors require oligomerization or association with ECM in order to function. Complex interactions among these factors and their receptors result in the

stimulation or inhibition of cell division, cell differentiation, cell signaling, and cell motility. Some factors act on their cell of origin (autocrine signaling); on neighboring cells (paracrine signaling); or on distant cells (endocrine signaling).

There are three broad classes of growth and differentiation factors. The first class includes the large polypeptide growth factors, e.g., epidermal growth factor, fibroblast growth factor, transforming growth factor, insulin-like growth factor, and platelet-derived growth factor. Each of these defines a family of related molecules which stimulate cell proliferation for wound healing, bone synthesis and remodeling, and regeneration of epithelial, epidermal, and connective tissues, and induce differentiation of embryonic tissues. Nerve growth factor functions specifically as a neurotrophic factor, and all induce differentiation of embryonic tissues. The second class includes the hematopoietic growth factors which stimulate the proliferation and differentiation of blood cells such as B-lymphocytes, T-lymphocytes, erythrocytes, platelets, eosinophils, basophils, neutrophils, macrophages, and their stem cell precursors. These factors include colony-stimulating factors, erythropoietin, and cytokines, e.g., interleukins, interferons (IFNs), and tumor necrosis factor (TNF). Cytokines are secreted by cells of the immune system and function in immunomodulation. The third class includes small peptide factors e.g., bombesin, vasopressin, oxytocin, endothelin, transferrin, angiotensin II, vasoactive intestinal peptide, and bradykinin, which function as hormones to regulate cellular functions other than proliferation.

Growth and differentiation factors have been shown to play critical roles in neoplastic transformation of cells in vitro and in tumor progression in vivo. Inappropriate expression of growth factors by tumor cells may contribute to vascularization and metastasis of melanotic tumors. In hematopoiesis, growth factor misregulation can result in anemias, leukemias and lymphomas. Certain growth factors, e.g., IFN, are cytotoxic to tumor cells both in vivo and in vitro. Moreover, growth factors and/or their receptors are related both structurally and functionally related to oncoproteins. In addition, growth factors affect transcriptional regulation of both proto-oncogenes and oncosuppressor genes. (Pimentel, E. (1994) Handbook of Growth Factors, CRC Press, Ann Arbor, MI, pp. 6-25.)

Proteolytic enzymes or proteases degrade proteins by reducing the activation energy needed for the hydrolysis of peptide bonds. The major families are the zinc, serine, cysteine,

thiol, and carboxyl proteases.

Zinc proteases, e.g., carboxypeptidase A, have a zinc ion bound to the active site, recognize C-terminal residues that contain an aromatic or bulky aliphatic side chain, and hydrolyze the peptide bond adjacent to the C-terminal residues. Serine proteases have an active site serine residue and include digestive enzymes, e.g., trypsin and chymotrypsin, components of the complement and blood-clotting cascades, and enzymes that control the degradation and turnover of extracellular matrix (ECM) molecules. Subfamilies of serine proteases include tryptases (cleavage after arginine or lysine), aspases (cleavage after aspartate), chymases (cleavage after phenylalanine or leucine), metases (cleavage after methionine), and serases (cleavage after serine). Cysteine proteases (e.g. cathepsin) are produced by monocytes, macrophages and other immune cells and are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Overproduction of these enzymes can cause the tissue destruction associated with rheumatoid arthritis and asthma. Thiol proteases, e.g., papain, contain an active site cysteine and are widely distributed within tissues. Thiol proteases effect catalysis through a thiol ester intermediate facilitated by a proximal histidine side chain. Carboxyl proteases, e.g., pepsin, are active only under acidic conditions (pH 2 to 3). The active site of pepsin contains two aspartate residues; when one aspartate is ionized and the other is not, the enzyme is active. A common feature of the carboxyl proteases is that they are inhibited by very low concentrations ( $10^{-10}$  M) of the inhibitor pepstatin. A substrate analog which induces structural changes at the active site of a protease functions as an antagonist or inhibitor.

Guanosine triphosphate-binding proteins (G proteins) participate in intracellular signal transduction and control regulatory pathways through cell surface receptors. These receptors respond to hormones, growth factors, neuromodulators, or other signaling molecules, by binding GTP. Binding of GTP leads to the production of cAMP which controls phosphorylation and activation of other proteins. During this process, the hydrolysis of GTP acts as an energy source as well as an on-off switch for the GTPase activity.

The G proteins are small proteins which consist of single 21-30 kDa polypeptides. They can be classified into five subfamilies: Ras, Rho, Ran, Rab, and ADP-ribosylation

factor. These proteins regulate cell growth, cell cycle control, protein secretion, and intracellular vesicle interaction. In particular, the Ras proteins are essential in transducing signals from receptor tyrosine kinases to serine/threonine kinases which control cell growth and differentiation. Mutant Ras proteins, which bind but can not hydrolyze GTP, are permanently activated and cause continuous cell proliferation or cancer.

All five subfamilies share common structural features and four conserved motifs, I to IV. Motif I is the most variable and has the signature of GXXXXGK, in which lysine interacts with the  $\beta$ - and  $\gamma$ -phosphate groups of GTP. Motif II, III, and IV have DTAGQE, NKXD, and EXSAX as their respective signatures and regulate the binding of  $\gamma$ -phosphate, GTP, and the guanine base of GTP, respectively. Most of the membrane-bound G proteins require a carboxy terminal isoprenyl group (CAAX), added posttranslationally, for membrane association and biological activity. The G proteins also have a variable effector region, located between motifs I and II, which is characterized as the interaction site for guanine nucleotide exchange factors or GTPase-activating proteins.

Eukaryotic cells are bound by a membrane and subdivided into membrane bound compartments. As membranes are impermeable to many ions and polar molecules, transport of these molecules is mediated by ion channels, ion pumps, transport proteins, or pumps. Symporters and antiporters regulate cytosolic pH by transporting ions and small molecules, e.g., amino acids, glucose, and drugs, across membranes; symporters transport small molecules and ions in the same direction, and antiporters, in the opposite direction. Transporter superfamilies include facilitative transporters and active ATP binding cassette transporters involved in multiple-drug resistance and the targeting of antigenic peptides to MHC Class I molecules. These transporters bind to a specific ion or other molecule and undergo conformational changes in order to transfer the ion or molecule across a membrane. Transport can occur by a passive, concentration-dependent mechanism or can be linked to an energy source such as ATP hydrolysis or an ion gradient.

Ion channels are formed by transmembrane proteins which form a lined passageway across the membrane through which water and ions, e.g.,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ , and  $\text{Cl}^-$ , enter and exit the cell. For example, chloride channels are involved in the regulation of the membrane electric potential as well as absorption and secretion of ions across the membrane. In



intracellular membranes of the Golgi apparatus and endocytic vesicles, chloride channels also regulate organelle pH. Electrophysiological and pharmacological studies suggest that a variety of chloride channels exist in different cell types and that many of these channels have one or more protein kinase phosphorylation sites.

5 Ion pumps are ATPases which actively maintain membrane gradients. Ion pumps can be grouped into three classes, e.g., P, V, and F, according to their structure and function. All have one or more binding sites for ATP on the cytosolic face of the membrane. The P-class ion pumps consist of two  $\alpha$  and two  $\beta$  transmembrane subunits, include  $\text{Ca}^{2+}$  ATPase and  $\text{Na}^+/\text{K}^+$  ATPase, and function in transporting  $\text{H}^+$ ,  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Ca}^{2+}$  ions. The V- and F-class  
10 ion pumps have similar structures, a cytosolic domain formed by at least five extrinsic polypeptides and at least 2 transmembrane proteins, and only transport  $\text{H}^+$ . F class  $\text{H}^+$  pumps have been identified from the membranes of mitochondria and chloroplast, and V-class  $\text{H}^+$  pumps regulate acidity inside lysosomes, endosomes, and plant vacuoles.

A family of structurally related intrinsic membrane proteins known as facilitative  
15 glucose transporters catalyze the movement of glucose and other selected sugars across the plasma membrane. The proteins in this family contain a highly conserved, large transmembrane domain made of 12 transmembrane  $\alpha$ -helices, and several less conserved, asymmetric, cytoplasmic and exoplasmic domains. (Pessin, J. E., and Bell, G.I. (1992) *Annu. Rev. Physiol.* 54:911-930.)

20 Amino acid transport is mediated by  $\text{Na}^+$  dependent amino acid transporters. These transporters are involved in gastrointestinal and renal uptake of dietary and cellular amino acids and the re-uptake of neurotransmitters. Transport of cationic amino acids is mediated by the system y<sup>+</sup> family members and the cationic amino acid transporter (CAT) family. Members of the CAT family share a high degree of sequence homology, and each contains  
25 12-14 putative transmembrane domains. (Ito, K. and Groudine, M. (1997) *J. Biol. Chem.* 272:26780-26786.)

Proton-coupled, 12 membrane-spanning domain transporters such as PEPT 1 and PEPT 2 are responsible for gastrointestinal absorption and for renal reabsorption of peptides using an electrochemical  $\text{H}^+$  gradient as the driving force. A heterodimeric peptide  
30 transporter, consisting of TAP 1 and TAP 2, is associated with antigen processing. Peptide

antigens are transported across the membrane of the endoplasmic reticulum so they can be presented to the major histocompatibility complex class I molecules. Each TAP protein consists of multiple hydrophobic membrane spanning segments and a highly conserved ATP-binding cassette. (Boll, M. et al. (1996) Proc. Natl. Acad. Sci. 93:284-289.)

5 Hormones are secreted molecules that circulate in the body fluids and bind to specific receptors on the surface of, or within, target tissue cells. Although they have diverse biochemical compositions and mechanisms of action, hormones can be grouped into two categories. One category consists of small lipophilic molecules that diffuse through the plasma membrane of target cells, bind to cytosolic or nuclear receptors, and form a complex  
10 alters gene expression. Examples of this category include retinoic acid, thyroxine, and the cholesterol derived steroid hormones, progesterone, estrogen, testosterone, cortisol, and aldosterone. These hormones have a long half-life, e.g., several hours to days, and long-term effects of their target cells. Their solubility in the blood may be increased by their association with carrier molecules. Within the target cell nucleus, hormone/receptor complexes bind to  
15 specific response elements in target gene regulatory regions.

A second category consists of hydrophilic hormones that function by binding to cell surface receptors and transducing the signal across the plasma membrane. Examples of this category include amino acid derivatives, such as catecholamines, e.g., epinephrine, norepinephrine, and histamine; peptide hormones, e.g., glucagon, insulin, gastrin, secretin,  
20 cholecystokinin, adrenocorticotrophic hormone, follicle stimulating hormone, luteinizing hormone, thyroid stimulating hormone, parathormone, and vasopressin. Peptide hormones are synthesized as inactive forms and stored in secretory vesicles. These hormones are activated by protease cleavage before being released from the cell. Many hydrophilic hormones have a very short half-life and effect, e.g., seconds to hours, and are inactivated by  
25 proteases in the blood. (Lodish et al. (1995) Molecular Cell Biology, Scientific American Books Inc., New York, NY, pp. 856-864.)

Neuropeptides and vasomediators (NP/VM) comprise a large family of endogenous signaling molecules. Included in the family are neurotransmitters such as bombesin, neuropeptide Y, neurotensin, neuromedin N, melanocortins, opioids, e.g., enkephalins,  
30 endorphins and dynorphins, galanin, somatostatin, tachykinins, vasopressin, and vasoactive

intestinal peptide, and circulatory system-borne signaling molecules, e.g., angiotensin, complement, calcitonin, endothelins, formyl-methionyl peptides, glucagon, cholecystokinin and gastrin. These proteins are synthesized as “pre-pro” molecules, and are activated and inactivated by proteolytic cleavage. NP/VMs can transduce signals directly, modulate the activity or release of other neurotransmitters and hormones, and act as catalytic enzymes in cascades. The effects of NP/VMs range from extremely brief or long-lasting (melanocortin-mediated changes in skin melanin). Regulatory molecules turn individual genes or groups of genes on and off in response to various inductive mechanisms of the cell or organism; act as transcription factors by determining whether or not transcription is initiated, enhanced, or repressed; and splice transcripts as dictated in a particular cell or tissue.

Although they interact with short stretches of DNA scattered throughout the entire genome, most gene expression is regulated near the site at which transcription starts or within the open reading frame of the gene being expressed. The regulated stretches of the DNA can be simple and interact with only a single protein, or they can require several proteins acting as part of a complex to regulate gene expression. The external features of the double helix which provide recognition sites are hydrogen bond donor and acceptor groups, hydrophobic patches, major and minor grooves, and regular, repeated stretches of sequences which cause distinct bends in the helix. The surface features of the regulatory molecule are complementary to those of the DNA.

Many of the transcription factors incorporate one of a set of DNA-binding structural motifs, each of which contains either  $\alpha$  helices or  $\beta$  sheets and binds to the major groove of DNA. Seven of the structural motifs common to transcription factors are helix-turn-helix, homeodomains, zinc finger, steroid receptor,  $\beta$  sheets, leucine zipper, and helix-loop-helix. (Pabo, C.O. and R.T. Sauer (1992) *Ann. Rev. Biochem.* 61:1053-95.) Other domains of transcription factors may form crucial contacts with the DNA. In addition, accessory proteins provide important interactions which may convert a particular protein complex to an activator or a repressor or may prevent binding. (Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing Co, New York, NY pp. 401-474.)

The discovery of new human signal peptide-containing proteins and the polynucleotides encoding these molecules satisfies a need in the art by providing new

compositions which are useful in the diagnosis, treatment, and prevention of cancer and immunological disorders.

# SUMMARY OF THE INVENTION

5 The invention features a substantially purified human signal peptide-containing protein (SIGP), having an amino acid sequence selected from the group consisting of SEQ ID NO:1 SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, 10 SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, 15 SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, 20 SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, and SEQ ID NO:77..

The invention further provides isolated and substantially purified polynucleotides encoding SIGP. In a particular aspect, the polynucleotide has a nucleic acid sequence selected from the group consisting of SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, 25 SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID 30 NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID

NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154.

In addition, the invention provides a polynucleotide, or fragment thereof, which hybridizes to any of the polynucleotides encoding an SIGP selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, and SEQ ID NO:77. In another aspect, the invention provides a composition comprising isolated and purified polynucleotides selected from the group consisting of SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID

NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154, or a fragment thereof.

The invention further provides a polynucleotide comprising the complement, or fragments thereof, of any one of the polynucleotides encoding SIGP. In another aspect, the invention provides compositions comprising isolated and purified polynucleotides comprising the complement of SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154, or fragments thereof.

The present invention further provides an expression vector containing at least a fragment of any one of the polynucleotides selected from the group consisting of SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154. In yet another aspect, the expression vector containing the polynucleotide is contained within a host cell.

The invention also provides a method for producing a polypeptide or a fragment thereof, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide encoding SIGP under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified SIGP in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to SIGP, as well as a purified agonist and a purified antagonist of SIGP.

The invention also provides a method for treating or preventing a cancer associated with the decreased expression or activity of SIGP, the method comprising the step of

administering to a subject in need of such treatment an effective amount of a pharmaceutical composition containing SIGP.

The invention also provides a method for treating or preventing a cancer associated with the increased expression or activity of SIGP, the method comprising the step of  
5 administering to a subject in need of such treatment an effective amount of an antagonist of SIGP.

The invention also provides a method for treating or preventing an immune response associated with the increased expression or activity of SIGP, the method comprising the step of administering to a subject in need of such treatment an effective amount of an antagonist of  
10 SIGP.

The invention also provides a method for detecting a nucleic acid sequence which encodes a human regulatory proteins in a biological sample, the method comprising the steps of: a) hybridizing a nucleic acid sequence of the biological sample to a polynucleotide sequence complementary to the polynucleotide encoding SIGP, thereby forming a  
15 hybridization complex; and b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the nucleic acid sequence encoding the human regulatory protein in the biological sample.

The invention also provides a microarray containing at least a fragment of at least one of the polynucleotides encoding a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID  
20 NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27,  
25 SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID  
30 NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60,



SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, and SEQ ID NO:77.

5 The invention also provides a method for detecting the expression level of a nucleic acid encoding a human regulatory protein in a biological sample, the method comprising the steps of hybridizing the nucleic acid sequence of the biological sample to a complementary polynucleotide, thereby forming hybridization complex; and determining expression of the nucleic acid sequence encoding a human regulatory protein in the biological sample by  
10 identifying the presence of the hybridization complex. In a preferred embodiment, prior to the hybridizing step, the nucleic acid sequences of the biological sample are amplified and labeled by the polymerase chain reaction.

## DESCRIPTION OF THE INVENTION

15 Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the  
20 appended claims.

It must be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to “a host cell” includes a plurality of such host cells, and a reference to “an antibody” is a reference to one or more antibodies and equivalents thereof  
25 known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices,  
30 and materials are now described. All publications mentioned herein are cited for the purpose

of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

5

# **DEFINITIONS**

"SIGP," as used herein, refers to the amino acid sequences of substantially purified SIGP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

10

The term "agonist," as used herein, refers to a molecule which, when bound to SIGP, increases or prolongs the duration of the effect of SIGP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of SIGP.

15

An "allele" or an "allelic sequence," as these terms are used herein, is an alternative form of the gene encoding SIGP. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

20

"Altered" nucleic acid sequences encoding SIGP, as described herein, include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same SIGP or a polypeptide with at least one functional characteristic of SIGP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding SIGP, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding SIGP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid

25

30

residues which produce a silent change and result in a functionally equivalent SIGP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of SIGP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments", "immunogenic fragments", or "antigenic fragments" refer to fragments of SIGP which are preferably about 5 to about 15 amino acids in length and which retain some biological activity or immunological activity of SIGP. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art. (See, e.g., Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5.)

The term "antagonist," as it is used herein, refers to a molecule which, when bound to SIGP, decreases the amount or the duration of the effect of the biological or immunological activity of SIGP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of SIGP.

As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fa, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind SIGP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing

antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to a specific nucleic acid sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic SIGP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity

exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence," as these terms are used herein, refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation, an aqueous solution, or a sterile composition. Compositions comprising polynucleotides encoding SIGP, e.g., SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154, or fragments thereof, may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS) and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

The phrase "consensus sequence," as used herein, refers to a nucleic acid sequence

which has been resequenced to resolve uncalled bases, extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GEL VIEW™ Fragment Assembly system (GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence .

As used herein, the term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding SIGP, by northern analysis is indicative of the presence of nucleic acids encoding SIGP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding SIGP.

The term "SIGP" refers to any or all of the human polypeptides, SIGP-1, SIGP-2, SIGP-3, SIGP-4, SIGP-5, SIGP-6, SIGP-7, SIGP-8, SIGP-9, SIGP-10, SIGP-11, SIGP-12, SIGP-13, SIGP-14, SIGP-15, SIGP-16, SIGP-17, SIGP-18, SIGP-19, SIGP-20, SIGP-21, SIGP-22, SIGP-23, SIGP-24, SIGP-25, SIGP-26, SIGP-27, SIGP-28, SIGP-29, SIGP-30, SIGP-31, SIGP-32, SIGP-33, SIGP-34, SIGP-35, SIGP-36, SIGP-37, SIGP-38, SIGP-39, SIGP-40, SIGP-41, SIGP-42, SIGP-43, SIGP-44, SIGP-45, SIGP-46, SIGP-47, SIGP-48, SIGP-49, SIGP-50, SIGP-51, SIGP-52, SIGP-53, SIGP-54, SIGP-55, SIGP-56, SIGP-57, SIGP-58, SIGP-59, SIGP-60, SIGP-61, SIGP-62, SIGP-63, SIGP-64, SIGP-65, SIGP-66, SIGP-67, SIGP-68, SIGP-69, SIGP-70, SIGP-71, SIGP-72, SIGP-73, SIGP-74, SIGP-75, SIGP-76, and SIGP-77.

A "deletion," as the term is used herein, refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative," as used herein, refers to the chemical modification of SIGP, of a polynucleotide sequence encoding SIGP, or of a polynucleotide sequence complementary to a polynucleotide sequence encoding SIGP. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is

one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "homology," as used herein, refers to a degree of complementarity. There may be partial homology or complete homology. The word "identity" may substitute for the word "homology." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% homology or identity). In the absence of non-specific binding, the substantially homologous sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MegAlign program (Lasergene software package, DNASTAR, Inc., Madison WI). The MegAlign program can create alignments between two or more sequences according to different methods, e.g., the Clustal Method. (Higgins, D.G. and Sharp, P.M. (1988) Gene 73:237-244.) The Clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no homology between the two amino acid

sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be calculated by the Clustal Method, or by other methods known in the art, such as the Jotun Hein Method. (See, e.g., Hein, J. (1990) Methods in Enzymology 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

“Human artificial chromosomes” (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355.)

The term “humanized antibody,” as used herein, refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization,” as the term is used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term “hybridization complex” as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words “insertion” or “addition,” as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

“Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term “microarray,” as used herein, refers to an array of distinct polynucleotides or



oligonucleotides arrayed on a substrate, such as paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The term "modulate," as it appears herein, refers to a change in the activity of SIGP. For example, modulation may cause an increase or a decrease in protein activity, binding  
5 characteristics, or any other biological, functional, or immunological properties of SIGP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to an oligonucleotide, nucleotide, polynucleotide, or any fragment thereof, to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like  
10 or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which are greater than about 60 nucleotides in length, and most preferably are at least about 100 nucleotides, at least about 1000 nucleotides, or at least about 10,000 nucleotides in length.

The terms "operably associated" or "operably linked," as used herein, refer to  
15 functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the encoded polypeptide but still bind to operator sequences that  
20 control expression of the polypeptide.

The term "oligonucleotide," as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. As used herein, the term "oligonucleotide" is substantially  
25 equivalent to the terms "amplimers," "primers," "oligomers," and "probes," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine  
30 confers solubility to the composition. PNAs preferentially bind complementary single

stranded DNA and RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell. (See, e.g., Nielsen, P.E. et al. (1993) *Anticancer Drug Des.* 8:53-63.)

The term "sample," as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding SIGP, or fragments thereof, or SIGP itself  
 5 may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

As used herein, the terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The  
 10 interaction is dependent upon the presence of a particular structure of the protein recognized by the binding molecule (i.e., the antigenic determinant or epitope). For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

As used herein, the term "stringent conditions" refers to conditions which permit  
 15 hybridization between polynucleotide sequences and the claimed polynucleotide sequences. Suitably stringent conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. In particular, stringency can be increased by  
 20 reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

For example, hybridization under high stringency conditions could occur in about 50% formamide at about 37°C to 42°C. Hybridization could occur under reduced stringency conditions in about 35% to 25% formamide at about 30°C to 35°C. In particular,  
 25 hybridization could occur under high stringency conditions at 42°C in 50% formamide, 5X SSPE, 0.3% SDS, and 200  $\mu$ g/ml sheared and denatured salmon sperm DNA. Hybridization could occur under reduced stringency conditions as described above, but in 35% formamide at a reduced temperature of 35°C. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the purine to pyrimidine ratio of the  
 30 nucleic acid of interest and adjusting the temperature accordingly. Variations on the above

ranges and conditions are well known in the art.

The term "substantially purified," as used herein, refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution," as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, and refers to cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of SIGP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

## THE INVENTION

The invention is based on the discovery of new human signal peptide-containing proteins, collectively referred to as SIGP and individually as SIGP-1, SIGP-2, SIGP-3,

SIGP-4, SIGP-5, SIGP-6, SIGP-7, SIGP-8, SIGP-9, SIGP-10, SIGP-11, SIGP-12, SIGP-13, SIGP-14, SIGP-15, SIGP-16, SIGP-17, SIGP-18, SIGP-19, SIGP-20, SIGP-21, SIGP-22, SIGP-23, SIGP-24, SIGP-25, SIGP-26, SIGP-27, SIGP-28, SIGP-29, SIGP-30, SIGP-31, SIGP-32, SIGP-33, SIGP-34, SIGP-35, SIGP-36, SIGP-37, SIGP-38, SIGP-39, SIGP-40, SIGP-41, SIGP-42, SIGP-43, SIGP-44, SIGP-45, SIGP-46, SIGP-47, SIGP-48, SIGP-49, SIGP-50, SIGP-51, SIGP-52, SIGP-53, SIGP-54, SIGP-55, SIGP-56, SIGP-57, SIGP-58, SIGP-59, SIGP-60, SIGP-61, SIGP-62, SIGP-63, SIGP-64, SIGP-65, SIGP-66, SIGP-67, SIGP-68, SIGP-69, SIGP-70, SIGP-71, SIGP-72, SIGP-73, SIGP-74, SIGP-75, SIGP-76, and SIGP-77; the polynucleotides encoding SIGP (SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154); and the use of these compositions for the diagnosis, treatment, or prevention of cancer and immunological disorders. Table 1 shows the sequence identification numbers, Incyte Clone identification number, cDNA library, NCBI sequence identifier and GenBank species description for each of the human signal peptide-containing proteins disclosed herein.

Nucleic acids encoding the SIGP-1 of the present invention were first identified in

TABLE 1

Protein	Nucleotide	Clone ID	Library	NCBI I.D.	Homolog species
SEQ ID NO:1	SEQ ID NO:78	305841	HEARNOT01	GI 505652	Homo sapiens
SEQ ID NO:2	SEQ ID NO:79	322866	EOSIHET02	GI 180141	Homo sapiens
SEQ ID NO:3	SEQ ID NO:80	546656	BEPINOT01	GI 2290530	Homo sapiens
SEQ ID NO:4	SEQ ID NO:81	693453	SYNORAT03	GI 1419461	Caenorhabditis elegans
SEQ ID NO:5	SEQ ID NO:82	866885	BRAITUT03	GI 1488683	Rattus norvegicus
SEQ ID NO:6	SEQ ID NO:83	124271	LUNGNOT03	GI 1523073	Homo sapiens
SEQ ID NO:7	SEQ ID NO:84	1255027	LUNGFET03	GI 1684845	Canis familiaris
SEQ ID NO:8	SEQ ID NO:85	1273453	TESTTUT02		
SEQ ID NO:9	SEQ ID NO:86	1275261	TESTTUT02	GI 56805	Rattus norvegicus
SEQ ID NO:10	SEQ ID NO:87	1281682	COLNNOT16		
SEQ ID NO:11	SEQ ID NO:88	1298305	BRSTNOT07		
SEQ ID NO:12	SEQ ID NO:89	1360501	LUNGNOT12	GI 1019433	Trypanosoma cruzi
SEQ ID NO:13	SEQ ID NO:90	1362406	LUNGNOT12	GI 2072705	Mycobacterium tuberculosis
SEQ ID NO:14	SEQ ID NO:91	1405329	LATRTUT02		
SEQ ID NO:15	SEQ ID NO:92	1415223	BRAINOT12	GI 205250	Rattus norvegicus
SEQ ID NO:16	SEQ ID NO:93	1416553	BRAINOT12		
SEQ ID NO:17	SEQ ID NO:94	1418517	KIDNNOT09		
SEQ ID NO:18	SEQ ID NO:95	1438165	PANCNOT08	GI 1515161	Caenorhabditis elegans
SEQ ID NO:19	SEQ ID NO:96	1440381	THYRNOT03	GI 1065459	Caenorhabditis elegans
SEQ ID NO:20	SEQ ID NO:97	1510839	LUNGNOT14	GI 2145052	Plasmodium berghei
SEQ ID NO:21	SEQ ID NO:98	1534876	SPLNNOT04		
SEQ ID NO:22	SEQ ID NO:99	1559131	SPLNNOT04	GI 496667	Saccharomyces cerevisiae
SEQ ID NO:23	SEQ ID NO:100	1601473	BLADNOT03		
SEQ ID NO:24	SEQ ID NO:101	1615809	BRAITUT12		
SEQ ID NO:25	SEQ ID NO:102	1634813	COLNNOT19	GI 2196924	Mus musculus
SEQ ID NO:26	SEQ ID NO:103	1638407	UTRSNOT06	GI 200547	Mus musculus

TABLE 1

Protein	Nucleotide	Clone ID	Library	NCBI I.D.	Homolog species
SEQ ID NO:27	SEQ ID NO:104	1653112	PROSTUT08	GI 49794	Mus musculus
SEQ ID NO:28	SEQ ID NO:105	1664634	BRSTNOT09	GI 1890375	Caenorhabditis elegans
SEQ ID NO:29	SEQ ID NO:106	1690990	PROSTUT10		
SEQ ID NO:30	SEQ ID NO:107	1704050	DUODNOT02	GI 1814277	Homo sapiens
SEQ ID NO:31	SEQ ID NO:108	1711840	PROSNOT16	GI 182651	Homo sapiens
SEQ ID NO:32	SEQ ID NO:109	1747327	STOMTUT02	GI 2062391	Homo sapiens
SEQ ID NO:33	SEQ ID NO:110	1750632	STOMTUT02	GI 459002	Caenorhabditis elegans
SEQ ID NO:34	SEQ ID NO:111	1812375	PROSTUT12		
SEQ ID NO:35	SEQ ID NO:112	1818761	PROSNOT20	GI 2493789	Homo sapiens
SEQ ID NO:36	SEQ ID NO:113	1824469	GBLATUT01	GI 2052134	Mycobacterium tuberculosis
SEQ ID NO:37	SEQ ID NO:114	1864292	PROSNOT19	GI 295671	Saccharomyces cerevisiae
SEQ ID NO:38	SEQ ID NO:115	1866437	THP1NOT01		
SEQ ID NO:39	SEQ ID NO:116	1871375	SKINBIT01		
SEQ ID NO:40	SEQ ID NO:117	1880830	LEUKNOT03	GI 1872521	Arabidopsis thaliana
SEQ ID NO:41	SEQ ID NO:118	1905325	OVARNOT07	GI 1754971	Homo sapiens
SEQ ID NO:42	SEQ ID NO:119	1919931	BRSTTUT01	GI 2104517	Homo sapiens
SEQ ID NO:43	SEQ ID NO:120	1969426	BRSTNOT04		
SEQ ID NO:44	SEQ ID NO:121	1969948	UCMCL5T01		
SEQ ID NO:45	SEQ ID NO:122	1988911	LUNGAST01	GI 56649	Rattus norvegicus
SEQ ID NO:46	SEQ ID NO:123	2061561	OVARNOT03		
SEQ ID NO:47	SEQ ID NO:124	2084489	PANCNOT04	GI 2262136	Arabidopsis thaliana
SEQ ID NO:48	SEQ ID NO:125	2203226	SPLNFET02	GI 1911776	Homo sapiens
SEQ ID NO:49	SEQ ID NO:126	2232884	PROSNOT16		
SEQ ID NO:50	SEQ ID NO:127	2328134	COLNNOT11	GI 1911776	Homo sapiens
SEQ ID NO:51	SEQ ID NO:128	2382718	ISLTNOT01	GI 1814277	Homo sapiens
SEQ ID NO:52	SEQ ID NO:129	2452208	ENDANOT01		

TABLE 1

Protein	Nucleotide	Clone ID	Library	NCBI I.D.	Homolog species
SEQ ID NO:53	SEQ ID NO:130	2457825	ENDANOT01	GI 1418625	Caenorhabditis elegans
SEQ ID NO:54	SEQ ID NO:131	2470740	THP1NOT03		
SEQ ID NO:55	SEQ ID NO:132	2479092	SMCANOT01		
SEQ ID NO:56	SEQ ID NO:133	2480544	SMCANOT01	GI 169345	Phaseolus vulgaris
SEQ ID NO:57	SEQ ID NO:134	2518547	BRAITUT21	GI 33969	Homo sapiens
SEQ ID NO:58	SEQ ID NO:135	2530650	GBLANOT02	GI 2204111	Bos taurus
SEQ ID NO:59	SEQ ID NO:136	2652271	THYMNOT04	GI 895855	Solanum lycopersicum
SEQ ID NO:60	SEQ ID NO:137	2746976	LUNGTUT11	GI 191983	Mus musculus
SEQ ID NO:61	SEQ ID NO:138	2753496	THP1AZS08	GI 987286	Schizosaccharomyces pombe
SEQ ID NO:62	SEQ ID NO:139	2781553	OVARUTUT03		
SEQ ID NO:63	SEQ ID NO:140	2821925	ADRETUT06		
SEQ ID NO:64	SEQ ID NO:141	2879068	UTRSTUT05	GI 870749	Homo sapiens
SEQ ID NO:65	SEQ ID NO:142	2886757	SINJNOT02	GI 1420026	Saccharomyces cerevisiae
SEQ ID NO:66	SEQ ID NO:143	2964329	SCORNOT04	GI 311667	Saccharomyces cerevisiae
SEQ ID NO:67	SEQ ID NO:144	2965248	SCORNOT04	GI 1478503	Homo sapiens
SEQ ID NO:68	SEQ ID NO:145	3000534	TYMNOT06	GI 1741868	Homo sapiens
SEQ ID NO:69	SEQ ID NO:146	3046870	HEAANOT01	GI 1067079	Caenorhabditis elegans
SEQ ID NO:70	SEQ ID NO:147	3057669	PONSAZT01	GI 260241	
SEQ ID NO:71	SEQ ID NO:148	3088178	HEAONOT03	GI 498997	Saccharomyces cerevisiae
SEQ ID NO:72	SEQ ID NO:149	3094321	BRSTNOT19	GI 793879	Saccharomyces cerevisiae
SEQ ID NO:73	SEQ ID NO:150	3115936	LUNGTUT13	GI 517174	Saccharomyces cerevisiae
SEQ ID NO:74	SEQ ID NO:151	3116522	LUNGTUT13	GI 1669560	Homo sapiens
SEQ ID NO:75	SEQ ID NO:152	3117184	LUNGTUT13	GI 1418628	Caenorhabditis elegans
SEQ ID NO:76	SEQ ID NO:153	3125156	LNODNOT05	GI 804750	Homo sapiens
SEQ ID NO:77	SEQ ID NO:154	3129120	LUNGTUT12	GI 1256890	Saccharomyces cerevisiae

Incyte Clone 305841 from the heart tissue cDNA library (HEARNOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:78, was derived from Incyte Clones 305841 (HEARNOT01), 22049 (ADENINB01), 168880 (LIVRNOT01), 1321915 (BLADNOT04), and the shotgun sequences SAWA02804, SAWA02781, SAWA01969, and SAWA01937.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1. SIGP-1 is 348 amino acids in length and has a potential amidation site at Q120; a potential N-glycosylation site at N181; two potential casein kinase II phosphorylation sites at S19 and T279; a potential glycosaminoglycan attachment site at S35; and three potential protein kinase C phosphorylation sites at S19, S268, and S343. SIGP-1 shares 56% identity with human GP36b glycoprotein (GI 505652). The fragment of SEQ ID NO:78 including the 5' region from about nucleotide 117 to about nucleotide 161 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, cardiovascular, hematopoietic and immune, and developmental cDNA libraries. Approximately 42% of these libraries are associated with neoplastic disorders, 28% with inflammation, and 21% with cell proliferation.

Nucleic acids encoding the SIGP-2 of the present invention were first identified in Incyte Clone 322866 from the eosinophil cDNA library (EOSIHET02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:79, was derived from Incyte Clones 322866 (EOSIHET02), 470107 (MMLR1DT01), 873933 (LUNGAST01), and 2268817 (UTRSNOT02)

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:2. SIGP-2 is 194 amino acids in length and has two potential N-glycosylation sites at N129 and N148; two potential casein kinase II phosphorylation sites at S74 and S151; four potential protein kinase C phosphorylation sites at S5, S74, S130, and S163; a potential tyrosine kinase phosphorylation site at Y171; two potential prokaryotic membrane lipoprotein lipid attachment sites at F15 and S61; and a transmembrane 4 protein family signature from G60 to L82. SIGP-2 shares 90% identity with CD53, a human cell surface antigen (GI 180141). The fragment of SEQ ID NO:79 from about nucleotide 624 to about nucleotide 686 is useful for hybridization. Northern analysis shows the expression of



this sequence in hematopoietic and immune, gastrointestinal, cardiovascular, reproductive, musculoskeletal, and neural cDNA libraries. Approximately 54% of these libraries are associated with inflammation, 39% with neoplastic disorders, and 11% with cell proliferation.

Nucleic acids encoding the SIGP-3 of the present invention were first identified in Incyte Clone 546656 from the bronchial epithelium primary cell line cDNA library (BEPINOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:80, was derived from Incyte Clones 546656 (BEPINOT01), 1316266 (BLADTUT02), 2095988 (BRAITUT02), 1318172 (BLADNOT04), 2809506 (TLYMNOT04), 1293412 and 1293630 (PGANNOT03), 2585048 (BRAITUT22), 2941370 (HEAONOT03), 2297230 (BRSTNOT05), 1233586 (LUNGFET03), and the shotgun sequence SAEA02986.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:3. SIGP-3 is 342 amino acids in length and has a potential amidation site at H4; a potential N-glycosylation site at N23; seven potential casein kinase II phosphorylation sites at S38, T90, T105, T124, S139, T284, and T324; three potential protein kinase C phosphorylation sites at S25, T71, and S200; two potential tyrosine kinase phosphorylation sites at Y13 and Y69; and a beta-transducin family Trp-Asp repeats signature sequence from I282 to I296. SIGP-3 shares 100% identity with human HAN11 (GI 2290530). The fragment of SEQ ID NO:80 from about nucleotide 107 to about nucleotide 139 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, cardiovascular, hematopoietic and immune, neural, urologic, and developmental cDNA libraries. Approximately 43% of these libraries are associated with neoplastic disorders, 25% with inflammation, and 20% with cell proliferation.

Nucleic acids encoding the SIGP-4 of the present invention were first identified in Incyte Clone 693453 from the synovial membrane cDNA library (SYNORAT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:81, was derived from Incyte Clones 693453 (SYNORAT03), 2505458 (CONUTUT01), 1527363 (UCMCL5T01), 1275308 (TESTTUT02), 1377126 (LUNGNOT10), 538256 (LNODNOT02), 3125441 (LNODNOT05), 1955296 (CONNNOT01), 1821536 (GBLATUT01), 2055631 (BEPINOT01), and 2028161 (KERANOT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:4. SIGP-4 is 656 amino acids in length and has a potential N-glycosylation site at N73; nine potential casein kinase II phosphorylation sites at S140, S191, T250, T252, S330, S340, S517, S617, and T630; a potential leucine zipper pattern from L430 to L451; four potential N-myristoylation sites at G77, G246, G484, and A651; eleven potential protein kinase C phosphorylation sites at S18, T90, S93, T318, S490, S503, S532, T565, T608, S609, and T629; and a potential tyrosine kinase phosphorylation site at Y326. SIGP-4 shares 20% identity with Caenorhabditis elegans protein encoded by T01G9.4 (GI 1419461). The fragment of SEQ ID NO:81 from about nucleotide 202 to about nucleotide 255 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, hematopoietic and immune, neural, and developmental cDNA libraries. Approximately 40% of these libraries are associated with neoplastic disorders, 30% with inflammation, and 30% with cell proliferation.

Nucleic acids encoding the SIGP-5 of the present invention were first identified in Incyte Clone 866885 from the brain tumor cDNA library (BRAITUT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:82, was derived from Incyte Clones 866885 (BRAITUT03), 2991983 (KIDNFET02), 067954 (HUVESTB01), and 1499109 (SINTBST01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:5. SIGP-5 is 236 amino acids in length and has a potential N-glycosylation site at N199; two potential casein kinase II phosphorylation sites at S8 and T72; a potential N-myristoylation site at G169; and three potential protein kinase C phosphorylation sites at T43, S96, and T201. SIGP-5 shares 24% identity with rat syntaxin (GI 1488683). The fragment of SEQ ID NO:82 from about nucleotide 43 to about nucleotide 93 is useful for hybridization. Northern analysis shows the expression of this sequence in hematopoietic and immune, reproductive, gastrointestinal, neural, cardiovascular, and developmental cDNA libraries. Approximately 43% of these libraries are associated with neoplastic disorders, 26% with inflammation, and 19% with cell proliferation.

Nucleic acids encoding the SIGP-6 of the present invention were first identified in Incyte Clone 1242271 from the lung tissue cDNA library (LUNGNOT03) using a computer

search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:83, was derived from Incyte Clones 1242271 (LUNGNOT03), 968114 (BRSTNOT05), 1251728 (LUNGFET03), and the shotgun sequence SAZA00142.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:6. SIGP-6 is 195 amino acids in length and has a potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S79; six potential casein kinase II phosphorylation sites at S79, T85, S113, T166, T171, and T188; three potential protein kinase C phosphorylation sites at S20, S150, and S185; and a potential mitochondrial energy transfer proteins signature from P25 to Y33. The fragment of SEQ ID NO:83 from about nucleotide 98 to about nucleotide 133 is useful for hybridization. Northern analysis shows the expression of this sequence in urologic, neural, reproductive, and cardiovascular cDNA libraries. Approximately 50% of these libraries are associated with neoplastic disorders, 14% with inflammation, and 21% with cell proliferation.

Nucleic acids encoding the SIGP-7 of the present invention were first identified in Incyte Clone 1255027 from the fetal lung cDNA library ( LUNGFET03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:84, was derived from Incyte Clones 1255027 (LUNGFET03), 2055704 (BEPINOT01), 1351096 (LATRTUT02), 835188 (PROSNOT07), and 1695810 (COLNNOT23).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:7. SIGP-7 is 608 amino acids in length and has a potential amidation site at T112; five potential N-glycosylation sites at N73, N110, N410, N436, and N478; two potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at S123 and S185; ten potential casein kinase II phosphorylation sites at T2, S75, S166, S170, S185, S274, S463, S505, S517, and T588; and thirteen potential protein kinase C phosphorylation sites at T19, S32, S46, T112, T221, S274, S299, T337, S373, S412, S431, S438, and S555. SIGP-7 shares 16% identity with canine pinin (GI 1684845). The fragment of SEQ ID NO:84 from about nucleotide 181 to about nucleotide 219 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal, neural, cardiovascular, and developmental cDNA libraries. Approximately 43% of these libraries are associated with neoplastic disorders, 21% with inflammation, and

20% with cell proliferation.

Nucleic acids encoding the SIGP-8 of the present invention were first identified in Incyte Clone 1273453 from the testicle cDNA library (TESTTUT02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:85, was derived from Incyte Clones 1273453 (TESTTUT02), 1970337 (UCMCL5T01), 1218926 (NEUTGMT01), 1881349 (LEUKNOT03), and 1722377 (BLADNT06).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:8. SIGP-8 is 267 amino acids in length and has a potential N glycosylation site at N230, five potential casein kinase II phosphorylation sites at S9, T45, T77, S190, and T263, and two potential protein kinase C phosphorylation sites at S232 and S236. The fragment of SEQ ID NO:85 from about nucleotide 140 to about nucleotide 175 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, cardiovascular, and hematopoietic and immune cDNA libraries. Approximately 42% of these libraries are associated with neoplastic disorders and 40% with immune response.

Nucleic acids encoding the SIGP-9 of the present invention were first identified in Incyte Clone 1275261 from the testicle cDNA library (TESTTUT02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:86, was derived from Incyte Clones 1275261 (TESTTUT02), 775078 (COLNNOT05), 514772 (MMLR1DT01), and 3224071 (COLNNON03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:9. SIGP-9 is 285 amino acids in length and has a potential amidation site at S260, three potential N glycosylation sites at N85, N100 and N156, a potential cAMP- and cGMP-dependent protein kinase phosphorylation site at T168, three potential casein kinase II phosphorylation sites at T168, T215, and S230, three potential protein kinase C phosphorylation sites at S163, S230, and S260, and a potential tyrosine kinase phosphorylation site at Y72. SIGP-9 shares 24% identity with rat OX-45 antigen preprotein (GI 56805). The fragment of SEQ ID NO:86 from about nucleotide 243 to about nucleotide 293 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal, and hematopoietic and immune cDNA

libraries. Approximately 50% of these libraries are associated with neoplastic disorders and 50% with immune response.

Nucleic acids encoding the SIGP-10 of the present invention were first identified in Incyte Clone 1281682 from the colon cDNA library (COLNNOT16) using a computer  
5 search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:87, was derived from Incyte Clones 2681940 (SINIUCT01), 1335652 (COLNNOT13), 2079572 (UTRSNOT08), 627405 (PGANNOT01) and 1281682 and 1282887 (COLNNOT16).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:10. SIGP-10 comprises a peptide of 76 amino acids in  
10 length, and has a potential signal peptide sequence from M1 to S18. The fragment of SEQ ID NO:87 encoding the potential signal peptide sequence from about nucleotide 908 through 970 is useful for hybridization. Northern analysis shows the expression of this sequence in gastrointestinal, neural, reproductive, and hematopoietic and immune cDNA libraries. Approximately 32% of these libraries are associated with neoplastic disorders and 53% with immune response.  
15

Nucleic acids encoding the SIGP-11 of the present invention were first identified in Incyte Clone 1298305 from the breast cDNA library (BRSTNOT09) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:88, was derived from Incyte Clones 1298305 (BRSTNOT09), 3451203 (UTRSNON03), 2529672  
20 (GBLAN0502), 2780863 (OVARTUT03), 927988 (BRAINOT04), 1684424 (PROSNOT15), 2243053 (PANCTUT02), and shotgun sequences SANA03310 and SANA00700.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:11. SIGP-11 is 147 amino acids in length and has a  
25 prokaryotic membrane lipoprotein lipid attachment site from L34 through C44. SIGP-11 also has a potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S91, and a potential protein kinase C phosphorylation site at S13. The fragment of SEQ ID NO:88 from about nucleotide 1561 to about nucleotide 1611 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal, and neural cDNA libraries. Approximately 50% of these libraries are associated with  
30

neoplastic disorders and 22% with immune response.

Nucleic acids encoding the SIGP-12 of the present invention were first identified in Incyte Clone 1360501 from the lung cDNA library (LUNGNOT12) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:89, was derived from Incyte Clones 1360501 (LUNGNOT12), 2121661 (BRSTNOT07), 1706518 (DUODNOT02) and shotgun sequences SAJA02519, SAJA00749, SAJA01160, and SANA00513.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:12. SIGP-12 is 261 amino acids in length and has six potential N glycosylation sites at N19, N28, N98, N104, N164 and N178. SIGP-12 also has five potential casein kinase II phosphorylation sites at T82, S83, T91, T160, and S233, and nine potential protein kinase C phosphorylation sites at T35, T60, T82, S121, S131, T184, S233, S237, and T242. SIGP-12 shares 22% identity with Trypanosoma cruzi mucin-like protein (GI 1019433). In addition, SIGP-12 shares two potential phosphorylation sites and a potential N-glycosylation site with the mucin-like protein. The fragment of SEQ ID NO:89 from about nucleotide 183 to about nucleotide 236 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, cardiovascular, and gastrointestinal cDNA libraries. Approximately 39% of these libraries are associated with neoplastic disorders and 26% with immune response.

Nucleic acids encoding the SIGP-13 of the present invention were first identified in Incyte Clone 1362406 from the lung cDNA library (LUNGNOT12) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:90, was derived from Incyte Clones 1362406 (LUNGNOT12), 1854401 (HNT3AZT01), 1570003 (UTRSNOT05) and shotgun sequences SANA03704, SANA00366, and SANA02152.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:13. SIGP-13 is 213 amino acids in length and has three potential protein kinase C phosphorylation sites at T40, S136, and T166. In addition, SIGP-13 has a highly hydrophobic signal peptide sequence from residue M1 to E34. SIGP-13 shares 20% identity with a Mycobacterium tuberculosis membrane protein (GI 2072705). The fragment of SEQ ID NO:90 encoding the potential signal peptide sequence

domain from about nucleotide 157 to about nucleotide 219 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, developmental, neural, and cardiovascular cDNA libraries. Approximately 50% of these libraries are associated with neoplastic disorders and 18% with immune response.

5 Nucleic acids encoding the SIGP-14 of the present invention were first identified in Incyte Clone 1405329 from the heart cDNA library (LATRTUT02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:91, was derived from Incyte Clones 1405329 (LATRTUT02), and 2830813 (TLYMNOT03).

10 In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:14. SIGP-14 is 67 amino acids in length and has a cell attachment sequence comprising R13 through D15. In addition, SIGP-14 has a potential casein kinase II phosphorylation site at T12, and a potential protein kinase C phosphorylation site at T42. The fragment of SEQ ID NO:91 from about nucleotide 36 to about nucleotide 95 is useful for hybridization. Northern analysis shows the expression of  
15 this sequence in cardiovascular, developmental, reproductive, and hematopoietic and immune cDNA libraries. Approximately 43% of these libraries are associated with neoplastic disorders and 21% with immune response.

20 Nucleic acids encoding the SIGP-15 of the present invention were first identified in Incyte Clone 1415223 from the brain cDNA library (BRAINOT12) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:92, was derived from Incyte Clones 1415223 (BRAINOT12) and 529786 (BRAINOT03).

25 In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:15. SIGP-15 is 161 amino acids in length and has a potential N-glycosylation site at N57, two potential casein kinase II phosphorylation sites at S84 and S96, and five potential protein kinase C phosphorylation sites at S11, T62, S75, S83, and S84. SIGP-15 shares 30% identity with rat Ly6C antigen (GI 205250). The fragment of SEQ ID NO:92 from about nucleotide 28 to about nucleotide 81 is useful for hybridization. Northern analysis shows the expression of this sequence in developmental, reproductive, and neural cDNA libraries. Approximately 33% of these libraries are associated with  
30 neoplastic disorders, 33% with cell proliferation, and 17% with immune response.

Nucleic acids encoding the SIGP-16 of the present invention were first identified in Incyte Clone 1416553 from the brain cDNA library (BRAINOT12) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:93, was derived from Incyte Clones 1416553 (BRAINOT12), 663124 (BRAINOT03) and shotgun sequences SANA01409, SANA03513, and SANA02713.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:16. SIGP-16 is 141 amino acids in length and has a glycosaminoglycan attachment site at S20. In addition, SIGP-16 has a potential casein kinase II phosphorylation site at S61, and a potential protein kinase C phosphorylation site at S53. The fragment of SEQ ID NO:93 from about nucleotide 784 to about nucleotide 831 is useful for hybridization. Northern analysis shows the expression of this sequence in neural cDNA libraries. Approximately 27% of these libraries are associated with neoplastic disorders, and 27% with neurological disorders.

Nucleic acids encoding the SIGP-17 of the present invention were first identified in Incyte Clone 1418517 from the kidney cDNA library (KIDNNOT09) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:94, was derived from Incyte Clones 1418517 (KIDNNOT09), 2456866 (ENDANOT01), 136927 (SYNORAB01), 1620442 (BRAITUT13), 1492394 (PROSNON01), 1534435 (SPLNNOT04), and 2505923 (CONUTUT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:17. SIGP-17 is 152 amino acids in length and has a potential N glycosylation site at N76; a potential cAMP- and cGMP-dependent protein kinase phosphorylation site at T67; four potential casein kinase II phosphorylation sites at S9, T30, S107, and S124; and three potential protein kinase C phosphorylation sites at T30, S34, and T78. The fragment of SEQ ID NO:94 from about nucleotide 49 to about nucleotide 99 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, cardiovascular, musculoskeletal, and gastrointestinal cDNA libraries. Approximately 44% of these libraries are associated with neoplastic disorders, 23% with immune response, and 20% with cell proliferation.

Nucleic acids encoding the SIGP-18 of the present invention were first identified in



Incyte Clone 1438165 from the pancreas cDNA library (PANCNOT08) using a computer search for amino acid alignments. A consensus sequence, SEQ ID NO:95, was derived from Incyte Clones 360389 (SYNORAB01), 485693 (HNT2RAT01), 1233177 (LUNGFET03), 1255551 (MENITUT03), 1438165 (PANCNOT08), 1554990 (BLADTUT04), and shotgun sequences SAOA00854 and SAOA00855.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:18. SIGP-18 is 742 amino acids in length and has a potential N-glycosylation site at N448; a microbodies C-terminal targeting signal in the triplet N740HL; twelve potential casein kinase II phosphorylation sites at S3, S53, S120, T122, T169, T178, S179, S195, T284, S290, S400, and S573; five potential protein kinase C phosphorylation sites at T178, S195, S208, S299, and S364; and two potential tyrosine kinase phosphorylation sites at Y296 and Y512. Cysteine residues, representing potential intramolecular disulfide bridging sites, are found at residues C87, C204, C312, C339, C343, C469, C497, C558, C657, C693, and C720. SIGP-18 shares 19% homology with C. elegans protein encoded by M163.4 (GI 1515161), including eight of the eleven cysteine residues found in SIGP-18. The fragment of SEQ ID NO:95 from about nucleotide 322 to about nucleotide 387 is useful for hybridization. Northern analysis shows the expression of this sequence in cardiovascular, male and female reproductive, and gastrointestinal cDNA libraries. Approximately 44% of these libraries are associated with neoplastic disorders, 23% with inflammation and the immune response, and 19% with fetal development.

Nucleic acids encoding the SIGP-19 of the present invention were first identified in Incyte Clone 1440381 from the thyroid cDNA library (THYRNOT03) using a computer search for amino acid alignments. A consensus sequence, SEQ ID NO:96, was derived from Incyte Clones 989671 (COLNNOT11), 1440381 (THYRNOT03), 3507668 (CONCNOT01), and shotgun sequences SAOA03364, SAOA02692, SAOA00489, SAOA02355, SAOA02405, SAOA01209, SAOA00809, and SAOA00274.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:19. SIGP-19 is 805 amino acids in length and has three potential N-glycosylation sites at N211, N215, and N327; one cAMP- and cGMP-dependent protein kinase potential phosphorylation sites at T749; sixteen potential casein kinase II

phosphorylation sites at S8, T54, T175, T228, S229, S250, S292, S329, T390, S401, S415, S471, S492, S671, T780, and S795; ten potential protein kinase C phosphorylation sites at S206, T396, S401, S442, T455, S600, S671, T683, S730, and S795; and two potential tyrosine kinase phosphorylation sites at Y437 and Y476. SIGP-19 shares 33% homology with a ubiquitin-conjugating, E2-like enzyme from *C. elegans* (GI 1065459). Both molecules share a "UBC domain" characteristic of ubiquitin-conjugating enzymes extending from approximately residue V559 to I647 of SIGP-19, and containing an active site cysteine residue, C614, required for thiolester formation. A characteristic proline-rich region, found at the N-terminal end of the UBC domain and extending from approximately P564 to P589 in SIGP-19, is also shared by both proteins. The fragment of SEQ ID NO:96 from about nucleotide 1678 to about nucleotide 1800 is useful for hybridization. Northern analysis shows the expression of this sequence in cardiovascular and male and female reproductive cDNA libraries. Approximately 50% of these libraries are associated with neoplastic disorders, 14% with inflammation and the immune response, and 19% with fetal development.

Nucleic acids encoding the SIGP-20 of the present invention were first identified in Incyte Clone 1510839 from the lung cDNA library (LUNGNOT14) using a computer search for amino acid alignments. A consensus sequence, SEQ ID NO:97, was derived from Incyte Clones 962326 (BRSTTUT03), 1383254 (BRAITUT08), 1510839 (LUNGNOT14), 1970949 (UCMCL5T01), 2214224 (SINTFET03), and shotgun sequences SAOA01059 and SAOA02595.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:20. SIGP-20 is 195 amino acids in length and has a potential signal peptide sequence between M1 and A39. SIGP-20 also has a potential N-glycosylation site at N83; and three potential casein kinase II phosphorylation sites at T161, T169, and T181; and three potential protein kinase C phosphorylation sites at T121, T143, and T153. SIGP-20 shares 21% homology with *Plasmodium berghei* merozoite surface protein-1 (GI 2145052). The fragment of SEQ ID NO:97 from about nucleotide 439 to about nucleotide 502 is useful for hybridization. Northern analysis shows the expression of this sequence in cardiovascular, male and female reproductive, and developmental cDNA libraries.

Approximately 48% of these libraries are associated with neoplastic disorders, 13% with inflammation and the immune response, and 19% with fetal development.

Nucleic acids encoding the SIGP-21 of the present invention were first identified in Incyte Clone 1534876 from the spleen cDNA library (SPLNNOT04) using a computer search for amino acid alignments. A consensus sequence, SEQ ID NO:98, was derived from Incyte Clones 1253004 (LUNGFET03), 1382838 (BRAITUT08), 1532501 (SPLNNOT04), 1534876 (SPLNNOT04), 1705806 (DUODNOT02), 1738301 (COLNNOT22), 1926209 (BRSTNOT02), and shotgun sequences SAOA00587, SAOA02048, and SAOA03535.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:21. SIGP-21 is 161 amino acids in length and has a potential signal peptide sequence between M1 and C13. SIGP-21 also has 17 cysteine residues with the potential for forming intramolecular disulfide bridges. Six of these cysteine residues, between residues C129 and C152, are found in a signature sequence for trypsin/alpha-amylase inhibitors that form a structure with intramolecular disulfide bridges. SIGP-21 has two potential casein kinase II phosphorylation sites at T25 and S35; and two potential protein kinase C phosphorylation sites at S35 and T87. The fragment of SEQ ID NO:98 from about nucleotide 406 to about nucleotide 477, which encompasses the trypsin/alpha-amylase inhibitor signature sequence, is useful for hybridization. Northern analysis shows the expression of this sequence in gastrointestinal and male and female reproductive cDNA libraries. Approximately 45% of these libraries are associated with neoplastic disorders and 28% with inflammation and the immune response.

Nucleic acids encoding the SIGP-22 of the present invention were first identified in Incyte Clone 1559131 from the spleen cDNA library (SPLNNOT04) using a computer search for amino acid alignments. A consensus sequence, SEQ ID NO:99, was derived from Incyte Clones 1559131 (SPLNNOT04), 1671080 (BMARNOT03), 1924001 (BRSTTUT01), and shotgun sequences SAPA01073 and SAOA02895.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:22. SIGP-22 is 160 amino acids in length and has cysteine residues capable of forming intramolecular disulfide bridges at C40, C47, C108, C114, C129, C154, and C158. SIGP-22 has one potential casein kinase II phosphorylation site at S9 and

one potential protein kinase C phosphorylation site at S31. SIGP-22 shares 26% homology with C-215 protein from Saccharomyces cerevisiae (GI 496667), including four of the cysteine residues found in SIGP-22. The fragment of SEQ ID NO:99 from about nucleotide 154 to about nucleotide 193 is useful for hybridization. Northern analysis shows the expression of this sequence in hematopoietic and male and female reproductive cDNA libraries. Approximately 33% of these libraries are associated with neoplastic disorders and 67% with the immune response.

Nucleic acids encoding the SIGP-23 of the present invention were first identified in Incyte Clone 1601473 from the bladder cDNA library (BLADNOT03) using a computer search for amino acid alignments. A consensus sequence, SEQ ID NO:100, was derived from Incyte Clones 1601473 (BLADNOT03), and shotgun sequences SAOA00407, SAOA02497, SAOA02747, and SAOA02958.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:23. SIGP-23 is 76 amino acids in length and has two cysteine residues with the potential of forming an intramolecular disulfide bridge at C58 and C72. SIGP-23 has one potential casein kinase II phosphorylation site at S7 and three potential protein kinase C phosphorylation sites at S7, T29, and T46. The fragment of SEQ ID NO:100 from about nucleotide 139 to about nucleotide 180 is useful for hybridization. Northern analysis shows the expression of this sequence in breast, brain, spleen, thyroid, and bladder cDNA libraries. Approximately 33% of these libraries are associated with neoplastic disorders, 17% with neural disorders, and 17% with immune disorders.

Nucleic acids encoding the SIGP-24 of the present invention were first identified in Incyte Clone 1615809 from the brain tumor cDNA library (BRAITUT12) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:101, was derived from Incyte Clones 1615809 (BRAITUT12), 924499 (BRAINOT04), 1273065 (TESTTUT02), 1517058 (PANCTUT01), 1596867 (BRAINOT14), and 1361446 (LUNGNOT12), and shotgun sequence SAOA02975.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:24. SIGP-24 is 336 amino acids in length and has 13 potential phosphorylation sites at T27, T72, S74, S76, T99, S104, S109, S140, S178, S210, T281,

S326, S39. SIGP-24 also has a potential signal peptide sequence between M1 and Y18. The fragment of SEQ ID NO:101 from about nucleotide 187 to about nucleotide 247 is useful for hybridization. Northern analysis shows the expression of this sequence in cardiovascular, gastrointestinal, neural, and reproductive cDNA libraries. Approximately 48% of these  
5 libraries are associated with neoplastic disorders and 21% with immune response.

Nucleic acids encoding the SIGP-25 of the present invention were first identified in Incyte Clone 1634813 from the cecal tissue cDNA library (COLNNOT19) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:102, was derived from Incyte Clones 1634813 (COLNNOT19), 2904583 (THYMNOT05), 1634813  
10 (COLNNOT19), and 1310492 (COLNFET02), and shotgun sequence SAPA04436.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:25. SIGP-25 is 150 amino acids in length and has one potential N-glycosylation site at N139; and five potential phosphorylation sites at T48, S118, S126, S135, and S136. SIGP-25 also has a potential signal peptide sequence encompassing residues  
15 M1-A23. SIGP-25 shares 28% identity with mouse beta chemokine, Exodus-2 (GI 2196924). The fragment of SEQ ID NO:102 from about nucleotide 175 to about nucleotide 235 is useful for hybridization. Northern analysis shows the expression of this sequence in gastrointestinal, developmental, hematopoietic, and immunological cDNA libraries. Approximately 50% of these libraries are associated with fetal development/cell proliferation and 25% with immune response.  
20

Nucleic acids encoding the SIGP-26 of the present invention were first identified in Incyte Clone 1638407 from the myometrial tissue cDNA library (UTRSNOT06) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:103, was derived from Incyte Clones 1638407 (UTRSNOT06), 3541410  
25 (SEMVNOT04), 1290413 (BRAINOT11), 1467841 (PANCTUT02), 1306495 (PLACNOT02), and 1907983 (CONNTUT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:26. SIGP-26 is 217 amino acids in length and has seven potential phosphorylation sites at T214, S68, S148, S189, S30, S110, and Y149. SIGP-26  
30 also has a potential signal peptide sequence between M1 and G31. SIGP-26 shares 18%

identity with a mouse proline-rich protein (GI 200547). The fragment of SEQ ID NO:103 from about nucleotide 146 to about nucleotide 206 is useful for hybridization. Northern analysis shows the expression of this sequence in gastrointestinal, hematopoietic, immunological, and reproductive cDNA libraries. Approximately 42% of these libraries are associated with neoplastic disorders and 39% with immune response.

Nucleic acids encoding the SIGP-27 of the present invention were first identified in Incyte Clone 1653112 from the prostate tumor tissue cDNA library (PROSTUT08) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:104, was derived from Incyte Clones 1653112 (PROSTUT08), 3450102 (UTRSNON03), 1969850 (UCMCL5T01), 1880259 (LEUKNOT03), 1504393 (BRAITUT07), and 394029 (TMLR2DT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:27. SIGP-27 is 504 amino acids in length and has eight potential phosphorylation sites at T338, T13, S38, T56, T132, T490, S33, and T472. SIGP-27 also has one potential leucine zipper pattern between L418 and L439. SIGP-27 shares 16% identity with mouse alpha-1 type-X collagen (GI 49794). The fragment of SEQ ID NO:104 from about nucleotide 130 to about nucleotide 190 is useful for hybridization. Northern analysis shows the expression of this sequence in cardiovascular, endocrine, hematopoietic, immunological, neural, and reproductive cDNA libraries. Approximately 55% of these libraries are associated with neoplastic disorders and 22% with immune response.

Nucleic acids encoding the SIGP-28 of the present invention were first identified in Incyte Clone 1664634 from the breast tissue cDNA library (BRSTNOT09) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:105, was derived from Incyte Clones 1664634 (BRSTNOT09) and 571656 (OVARNON01), and shotgun sequences SAPA04612, SAPA00377, and SAPA03034.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:28. SIGP-28 is 320 amino acids in length and has two potential N-glycosylation sites at N122 and N139; and eight potential phosphorylation sites at T30, S52, S109, S162, S220, S96, T258, and S280. SIGP-28 also has a potential signal peptide

sequence between M1 and A21. SIGP-28 shares 28% identity with a C. elegans protein encoded by F32A7.4 (GI 1890375). The fragment of SEQ ID NO:105 from about nucleotide 280 to about nucleotide 340 is useful for hybridization. Northern analysis shows the expression of this sequence in cardiovascular, gastrointestinal, hematopoietic, immunological, neural, and reproductive cDNA libraries. Approximately 38% of these libraries are associated with neoplastic disorders and 32% with immune response.

Nucleic acids encoding the SIGP-29 of the present invention were first identified in Incyte Clone 1690990 from the prostatic tumor tissue cDNA library (PROSTUT10) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:106, was derived from Incyte Clone 1690990 (PROSTUT10), and shotgun sequences SAPA01051, SAPA04063, SAPA01670, SAPA02170, SAPA01946, and SAPA00282.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:29. SIGP-29 is 117 amino acids in length and has one potential N-glycosylation site at N96; four potential phosphorylation sites at S16, S34, T78, and S62; and one potential N-myristoylation site at G5. SIGP-29 also has one potential microbodies C-terminal targeting signal at S115. The fragment of SEQ ID NO:106 from about nucleotide 1000 to about nucleotide 1062 is useful for hybridization. Northern analysis shows the expression of this sequence in gastrointestinal, reproductive, dermal, musculoskeletal, neural, and urogenital cDNA libraries. Approximately 77% of these libraries are associated with neoplastic disorders and 8% with immune response.

Nucleic acids encoding the SIGP-30 of the present invention were first identified in Incyte Clone 1704050 from the duodenal cDNA library (DUODNOT02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:107, was derived from Incyte Clones 865233 (BRAITUT03), 1359660 (LUNGNOT12), and 1704050 (DUODNOT02) and shotgun sequence SAPA02672.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:30. SIGP-30 is 298 amino acids in length and has one potential amidation site at P226; four potential N-glycosylation sites at N98, N187, N236, and N277; seven potential casein kinase II phosphorylation sites at T39, S59, T100, T149, S205, T284, and S286; three potential protein kinase C phosphorylation sites at T52, S58,

and S279; a potential signal sequence from M1 to G22; and a potential transmembrane spanning region from M230 to A261. SIGP-30 contains two potential immunoglobulin superfamily domains, from about F29 to about L131 and from about S138 to about R224. SIGP-30 shares 25% identity with the human A33 antigen precursor expressed in normal human colonic and small bowel epithelium and in human colon cancers (GI 1814277). In addition, the position of the hydrophobic transmembrane domain is conserved between these molecules. The cysteine residues at C50, C109, C139, C155, C214, and C254 are conserved between these molecules. The fragment of SEQ ID NO:107 from about nucleotide 1150 to about nucleotide 1209 is useful for hybridization. Northern analysis shows the expression of this sequence in neural, reproductive, cardiovascular, and endocrine cDNA libraries. Approximately 68% of these libraries are associated with cancer and 9% with immune response.

Nucleic acids encoding the SIGP-31 of the present invention were first identified in Incyte Clone 1711840 from the prostate cDNA library (PROSNOT16) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:108, was derived from Incyte Clones 1711840 (PROSNOT16) and 2550483 (LUNGTUT06) and shotgun sequence SAQA03185.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:31. SIGP-31 is 118 amino acids in length and has three potential protein kinase C phosphorylation sites at S48, T103, and S109; and a potential signal peptide sequence from M1 to A20. SIGP-31 shares 61% identity with human midkine, a retinoic acid-responsive heparin binding factor involved in regulation of growth and differentiation (GI 182651). The fragment of SEQ ID NO:108 from about nucleotide 511 to about nucleotide 555 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal, developmental, neural, and cardiovascular cDNA libraries. Approximately 58% of these libraries are associated with cancer, 16% with immune response, and 23% with fetal/proliferating cells.

Nucleic acids encoding the SIGP-32 of the present invention were first identified in Incyte Clone 1747327 from the stomach tumor cDNA library (STOMTUT02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID



NO:109, was derived from Incyte Clones 475228 (MMLR2DT01), 1500771 (SINTBST01), 1880656 (LEUKNOT03), 1747327 (STOMTUT02), and 2720285 (LUNGTUT10).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:32. SIGP-32 is 248 amino acids in length and has one potential N-glycosylation site at N56; three potential casein kinase II phosphorylation sites at S46, S134, and S140; and one potential protein kinase C phosphorylation site at T217. SIGP-32 shares 100% identity with human K12 protein precursor which is expressed in breast cancer cells and peripheral blood leukocytes (GI 2062391). Northern analysis shows the expression of this sequence in gastrointestinal, reproductive, hematopoietic/immune, and cardiovascular cDNA libraries. Approximately 59% of these libraries are associated with cancer and 35% with immune response.

Nucleic acids encoding the SIGP-33 of the present invention were first identified in Incyte Clone 1750632 from the stomach tumor cDNA library (STOMTUT02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:110, was derived from Incyte Clones 1521122 (BLADTUT04) and 1750632 (STOMTUT02) and shotgun sequences SAEA02182 and SAEA10021.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:33. SIGP-33 is 150 amino acids in length and has one potential protein kinase C phosphorylation site at S6. SIGP-33 shares 49% identity with the *C. elegans* protein encoded by R151.6 (GI 459002). The fragment of SEQ ID NO:110 from about nucleotide 514 to about nucleotide 573 is useful for hybridization. Northern analysis shows the expression of this sequence in cardiovascular and gastrointestinal cDNA libraries. Approximately 88% of these libraries are associated with cancer and 13% with immune response.

Nucleic acids encoding the SIGP-34 of the present invention were first identified in Incyte Clone 1812375 from the prostate tumor cDNA library (PROSTUT12) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:111, was derived from Incyte Clones 775001 (COLNNOT05), 834305 (PROSNOT07), 1504623 (BRAITUT07), and 1812375 (PROSTUT12) and shotgun sequences SAQA02414,

SATA00657, and SATA01478.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:34. SIGP-34 is 431 amino acids in length and has four potential N-glycosylation sites at N11, N49, N73, and N312; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S197; six potential casein kinase II phosphorylation sites at T38, S79, S130, S165, S177, and T188; three potential protein kinase C phosphorylation sites at S184, T254, and S337; and a potential high affinity calcium ion-binding, vitamin K-dependent carboxylation domain between W371 and W408. The fragments of SEQ ID NO:111 from about nucleotide 222 to about nucleotide 282 and the potential carboxylation domain encoded from about nucleotide 1267 to about nucleotide 1380 are useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, gastrointestinal, cardiovascular, and hematopoietic/immune DNA libraries. Approximately 52% of these libraries are associated with cancer, 24% with immune response, and 20% with fetal/proliferating cells.

Nucleic acids encoding the SIGP-35 of the present invention were first identified in Incyte Clone 1818761 from the prostate cDNA library (PROSNOT20) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:112, was derived from Incyte Clone 1818761 (PROSNOT20) and shotgun sequences SAJA00040, SAJA00601, SAJA01791, and SAJA02873.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:35. SIGP-35 is 278 amino acids in length and has one potential N-glycosylation site at N91; three potential casein kinase II phosphorylation sites at S9, S125, and S156; two potential protein kinase C phosphorylation sites at S77 and S224; one potential tyrosine kinase phosphorylation site at Y258; and a potential signal sequence from M1 to A30. SIGP-35 has fourteen consecutive collagen repeats (G-X-P or G-X-X) from G97 to P138 which could form a triple helical structure. SIGP-35 shares 28% identity with the human adipocyte complement-related protein precursor (Acrp30) (GI 2493789). The fragment of SEQ ID NO:112 from about nucleotide 157 to about nucleotide 210 is useful for hybridization. Northern analysis shows the expression of this sequence in developmental, dermal, gastrointestinal, hematopoietic/immune, neural, and

reproductive cDNA libraries. Approximately 29% of these libraries are associated with cancer, 43% with immune response, and 29% with fetal development.

Nucleic acids encoding the SIGP-36 of the present invention were first identified in Incyte Clone 1824469 from the gallbladder tumor cDNA library (GBLADTUT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:113, was derived from Incyte Clones 1664262 (BRSTNOT09), 1733422 (BRSTTUT08), 1824469 (GBLADTUT01), 2057044 (BEPINOT01), and 2449822 (ENDANOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:36. SIGP-36 is 286 amino acids in length and has one potential N-glycosylation site at N271; four potential casein kinase II phosphorylation sites at S50, S192, T230, and T251; and five potential protein kinase C phosphorylation sites at T29, T41, S50, T160, and T273. SIGP-36 shares 24% identity with the Mycobacterium tuberculosis protein encoded by MTCI237.14c (GI 2052134). The fragment of SEQ ID NO:113 from about nucleotide 415 to about nucleotide 468 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal, hematopoietic/immune, and neural cDNA libraries. Approximately 49% of these libraries are associated with cancer, 21% with immune response, and 21% with fetal/proliferating cells.

Nucleic acids encoding the SIGP-37 of the present invention were first identified in Incyte Clone 1864292 from the diseased prostate cDNA library (PROSNOT19) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:114, was derived from Incyte Clone 1864292 (PROSNOT19) and shotgun sequences SARA02195, SARA03070, SARA03675, and SATA02454.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:37. SIGP-37 is 404 amino acids in length and has one potential amidation site at V136; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S66; twenty potential casein kinase II phosphorylation sites at S23, T27, T74, S110, S111, S118, T122, S143, S145, S205, S207, S218, S219, S220, T252, S254,

S328, S330, S385, and T393; and twelve potential protein kinase C phosphorylation sites at T27, S76, T81, S140, S161, S176, S229, T285, S309, S356, S367, and S398. SIGP-37 shares 18% identity with the S. cerevisiae protein encoded by SRP40, a weak suppressor of a mutant of the subunit AC40 of DNA-dependent RNA polymerases I and II (GI 295671).

- 5 The fragment of SEQ ID NO:114 from about nucleotide 193 to about nucleotide 222 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, cardiovascular, and hematopoietic/immune cDNA libraries. Approximately 75% of these libraries are associated with cancer and 25% with immune response.

10 Nucleic acids encoding the SIGP-38 of the present invention were first identified in Incyte Clone 1866437 from the human promonocyte cell line cDNA library (THP1NOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:115, was derived from Incyte Clones 817970 (OVRTUT01), 825684 (PROSNOT06), 1866437 (THP1NOT01), 2190170 (PROSNOT26), and 3137972 (SMCCNOT02).

15 In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:38. SIGP-38 is 405 amino acids in length and has one potential N-glycosylation site at N378; one potential cAMP- and cGMP-phosphorylation site at S332; nine potential casein kinase II phosphorylation sites at T34, S51, T77, S107, S158, S264, T266, S296, and S332; and one potential protein kinase C phosphorylation site at S68. The fragment of SEQ ID NO:115 from about nucleotide 85 to about nucleotide 144 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, hematopoietic/immune, neural, and developmental cDNA libraries. Approximately 37% of these libraries are associated with cancer, 33% with immune response, and 22% with fetal/proliferating cells.

20  
25 Nucleic acids encoding the SIGP-39 of the present invention were first identified in Incyte Clone 1871375 from the leg skin erythema nodosum cDNA library (SKINBIT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:116, was derived from Incyte Clones 1428052 (SINTBST01), 1871375 (SKINBIT01), and 3210563 (BLADNOT08).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:39. SIGP-39 is 177 amino acids in length and has one potential casein kinase II phosphorylation site at S133; one potential glycosaminoglycan attachment site at S28GGG; and four potential protein kinase C phosphorylation sites at S44, S82, S115, and T148. SIGP-39 contains a signature sequence shared by the binding domains of receptors for lymphokines, hematopoietic growth factors and growth hormone-related molecules at S52RWSLWS. The fragment of SEQ ID NO:116 encoding the sequence surrounding the receptor binding domain signature from about nucleotide 190 to about nucleotide 249 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, cardiovascular, gastrointestinal, and developmental cDNA libraries. Approximately 44% of these libraries are associated with cancer and 19% with immune response.

Nucleic acids encoding the SIGP-40 of the present invention were first identified in Incyte Clone 1880830 from the leukocyte cDNA library (LEUKNOT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:117, was derived from Incyte Clones 361577 (PROSNOT01); 2113591 (BRAITUT03); 1880830 (LEUKNOT03) and shotgun sequences SATA03292 and SATA00377.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:40. SIGP-40 is 197 amino acids in length and has a potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S121; and four potential protein kinase C phosphorylation sites at T3, S57, T107, and T153. SIGP-40 shares 15% identity with the Arabidopsis thaliana zinc-finger protein Lsd1 (GI 1872521). The fragment of SEQ ID NO:117 from about nucleotide 567 to about nucleotide 621 is useful for hybridization. Northern analysis shows the expression of this sequence in neural and reproductive cDNA libraries. Approximately 49% of these libraries are associated with neoplastic disorders, 24% with immune response, and 16% with fetal development.

Nucleic acids encoding the SIGP-41 of the present invention were first identified in Incyte Clone 1905325 from the ovary cDNA library (OVARNOT07) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:118, was

derived from Incyte Clones 1905325 (OVARNOT07); 621454 (PGANNOT01); 621326 (PGANNOT01); 1264490 (SYNORAT05); 487357 (HNT2AGT01); 773311 (COLNCRT01); and shotgun sequence SATA03582.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:41. SIGP-41 is 302 amino acids in length and has two potential N-glycosylation sites at N80 and N252; three potential casein kinase II phosphorylation sites at S46, T58, and S143; and four potential protein kinase C phosphorylation sites at T58, S62, T147, and S300. SIGP-41 shares 27% identity with human necdin-related protein (GI 1754971). The fragment of SEQ ID NO:118 from about nucleotide 1701 to about nucleotide 1800 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, and gastrointestinal cDNA libraries. Approximately 51% of these libraries are associated with neoplastic disorders and 20% with immune response, and 18% with fetal development.

Nucleic acids encoding the SIGP-42 of the present invention were first identified in Incyte Clone 1919931 from the breast tumor cDNA library (BRSTTUT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:119, was derived from Incyte Clones 1919931 (BRSTTUT01) and shotgun sequences SATA02529, SATA01526 and SATA00892.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:42. SIGP-42 is 164 amino acids in length and has one potential casein kinase II phosphorylation site at T68; and two potential protein kinase C phosphorylation sites at T81 and S85. SIGP-42 shares 12% identity with human chemokine receptor (GI 2104517). The fragment of SEQ ID NO:119 from about nucleotide 585 to about nucleotide 630 is useful for hybridization. Northern analysis shows the expression of this sequence in hematopoietic/immune, reproductive, and neural cDNA libraries. Approximately 50% of these libraries are associated with neoplastic disorders and 38% with immune response.

Nucleic acids encoding the SIGP-43 of the present invention were first identified in Incyte Clone 1969426 from the breast tissue cDNA library (BRSTNOT04) using a

computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:120, was derived from Incyte Clones 1969426 (BRSTNOT04), 2373191 (ADRENOT07), 1225516 (COLNTUT02), 1555912 (BLADTUT04), 1449240 (PLACNOT02), and shotgun sequences SAZA01457 and SAZA00207.

5 In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:43. SIGP-43 is 235 amino acids in length and has one potential N-glycosylation site at N146; one potential glycosaminoglycan attachment site at S82; and four potential protein kinase C phosphorylation sites at T16, T43, S228, and S231. The fragment of SEQ ID NO:120 from about nucleotide 243 to about nucleotide 282 is  
10 useful for hybridization. Northern analysis shows the expression of this sequence in neural, reproductive, hematopoietic/immune, cardiovascular, gastrointestinal, and muscle cDNA libraries. Approximately 46% of these libraries are associated with neoplastic disorders and 28% with immune response.

Nucleic acids encoding the SIGP-44 of the present invention were first identified in  
15 Incyte Clone 1969948 from the umbilical cord cDNA library (UCMCL5T01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:121, was derived from Incyte Clones 1969948 (UCMCL5T01) and shotgun sequences SATA01513 and SATA00507.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:44. SIGP-44 is 203 amino acids in length and has three  
20 potential casein kinase II phosphorylation sites at T23, S114, and S120; one potential protein kinase C phosphorylation site at T105; and one potential tyrosine kinase phosphorylation site at Y47. The fragment of SEQ ID NO:121 from about nucleotide 162 to about nucleotide 216 is useful for hybridization. Northern analysis shows the expression of this sequence in  
25 gastrointestinal, hematopoietic/immune, reproductive, and cardiovascular cDNA libraries. Approximately 35% of these libraries are associated with neoplastic disorders and 24% with immune response.

Nucleic acids encoding the SIGP-45 of the present invention were first identified in Incyte Clone 1988911 from the lung cDNA library (LUNGAST01) using a computer

search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:122, was derived from Incyte Clones 1988911 (LUNGAST01), 860576 (BRAITUT03), 3188894 (THYMNON04), 1466606 (PANCTUT02), 1920945 (BRSTTUT01), 1502970 (BRAITUT07), and shotgun sequence SAZC00040.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:45. SIGP-45 is 359 amino acids in length and has nine potential casein kinase II phosphorylation sites at S34, S47, S115, T120, T141, S157, S182, S214, and S331; three potential protein kinase C phosphorylation sites at S34, T259, and S325; and one potential tyrosine kinase phosphorylation site at Y241. SIGP-45 shares 16% identity with rat myosin heavy chain (GI 56649). The fragment of SEQ ID NO:122 from about nucleotide 477 to about nucleotide 558 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, hematopoietic/immune, gastrointestinal, and cardiovascular cDNA libraries. Approximately 47% of these libraries are associated with neoplastic disorders, 33% with immune response, and 20% with fetal development.

Nucleic acids encoding the SIGP-46 of the present invention were first identified in Incyte Clone 2061561 from the ovary cDNA library (OVARNOT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:123, was derived from Incyte Clones 2061561 (OVARNOT03), 2208104 (SINTFET03), 2058750 (OVARNOT03), and shotgun sequences SAZA00915, SAZA00150, and SAZA00799.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:46. SIGP-46 is 150 amino acids in length and has two potential amidation sites at F57 and W74; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at T62; two potential casein kinase II phosphorylation sites at T101 and T110; and two potential protein kinase C phosphorylation sites at T28 and T97. The fragment of SEQ ID NO:123 from about nucleotide 82 to about nucleotide 168 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, gastrointestinal, and cardiovascular cDNA libraries. Approximately 54% of these libraries are associated with neoplastic disorders and 22% with immune response.



Nucleic acids encoding the SIGP-47 of the present invention were first identified in Incyte Clone 2084489 from the pancreas cDNA library (PANCNOT04) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:124, was derived from Incyte Clones 2084489 (PANCNOT04) and shotgun sequences SAJA00837,  
5 SAJA00793, SAJA01402, SAJA01533, and SAJA01490.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:47. SIGP-47 is 402 amino acids in length and has one potential N-glycosylation site at N191; seven potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at S22, S23, T80, S81, S202, S248, and S382; twenty-two  
10 potential casein kinase II phosphorylation sites at S8, S35, S56, S107, T152, S166, S170, S202, S206, S208, T212, S214, S216, T244, S252, S256, T264, T287, S288, T327, S362, S387; ten potential protein kinase C phosphorylation sites at S16, S116, S140, T180, S193, S194, T236, T244, S252, and S387; and one potential tyrosine kinase phosphorylation site at Y361. SIGP-47 shares 28% identity with an A. thaliana protein of unknown function (GI  
15 2262136). The most conserved region, residues 296 to 386 of SIGP-47, shares 70% identity with residues 299 to 386 of the A. thaliana protein. In addition, the potential amidation site at A314 in SIGP-47 is conserved as one potential amidation site at Q317 in the A. thaliana protein; and four potential protein kinase C or cAMP- and cGMP dependent protein kinase phosphorylation sites at S193, T236, S252 and Y361 in SIGP-47 are  
20 conserved as potential phosphorylation sites at S165, S219, T247, and Y364 respectively in the A. thaliana protein. The fragment of SEQ ID NO:124 from about nucleotide 468 to about nucleotide 531 is useful for hybridization. Northern analysis shows the expression of this sequence in neural, gastrointestinal and cardiovascular cDNA libraries. Approximately 50% of these libraries are associated with neoplastic disorders and 20%  
25 with trauma.

Nucleic acids encoding the SIGP-48 of the present invention were first identified in Incyte Clone 2203226 from the fetal spleen cDNA library (SPLNFET02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:125, was derived from Incyte Clones 2203226 (SPLNFET02), 2215960 (SINTFET03), 1291348

(BRAINOT11), 1874915 (LEUKNOT02), and 275828 (TESTNOT03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:48. SIGP-48 is 311 amino acids in length and has one potential amidation site at V117; one potential casein kinase II phosphorylation site at T215; and three potential protein kinase C phosphorylation sites at T13, S18, and T263. SIGP-48 shares 32% identity with a human putative Rab5 interacting protein (GI 1911776). The fragment of SEQ ID NO:125 from about nucleotide 747 to about nucleotide 846 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, cardiovascular, neural, and gastrointestinal cDNA libraries. Approximately 44% of these libraries are associated with neoplastic disorders, 30% with fetal/proliferative cells and tissues, and 23% with immune response.

Nucleic acids encoding the SIGP-49 of the present invention were first identified in Incyte Clone 2232884 from the prostate cDNA library (PROSNOT16) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:126, was derived from Incyte Clones 2232884 (PROSNOT16), 2728528 (OVARTUT05), 2232884 (PROSNOT16), and shotgun sequences SASA00238 and SASA00455.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:49. SIGP-49 is 316 amino acids in length and has one potential N-glycosylation site at N140; five potential casein kinase II phosphorylation sites at S3, T8, S29, S85, and T198; and two potential protein kinase C phosphorylation sites at T28 and S60. The fragment of SEQ ID NO:126 from about nucleotide 180 to about nucleotide 279 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, urologic, and neural cDNA libraries. Approximately 77% of these libraries are associated with neoplastic disorders.

Nucleic acids encoding the SIGP-50 of the present invention were first identified in Incyte Clone 2328134 from the colon cDNA library (COLNNOT11) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:127, was derived from Incyte Clones 2328134 (COLNNOT11), 1870180 (SKINBIT01), 081403 (SYNORAB01), and 851547 (NGANNOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:50. SIGP-50 is 346 amino acids in length and has two potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at residues S43 and S217; one potential casein kinase II phosphorylation site at residue T96; and five potential protein kinase C phosphorylation sites at residues T2, T15, T39, T247, and S301. SIGP-50 shares 33% identity with the human putative rab5-interacting protein (GI 1911776) and the casein kinase II phosphorylation site at residue T96. The fragment of SEQ ID NO:127 encoding the potential extracellular ligand binding domain from about nucleotide 16 to about nucleotide 76 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal, cardiovascular, and neural cDNA libraries. Approximately 44% of these libraries are associated with cancer, 28% are associated with immune response, and 20% with fetal disorders.

Nucleic acids encoding the SIGP-51 of the present invention were first identified in Incyte Clone 2382718 from the pancreatic cDNA library (ISLTNOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:128, was derived from Incyte Clones 2382718 (ISLTNOT01), 3472492 (LUNGNOT27), 014756 (THP1PLB01), 1731885 (BRSTTUT08), 1889866 (BLADTUT07), and 1447744 (PLACNOT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:51. SIGP-51 is 299 amino acids in length and has one potential N-glycosylation site at residue N185; one cAMP- and cGMP-dependent protein kinase phosphorylation site at T273; nine potential casein kinase II phosphorylation sites at S34, S82, T100, S118, T152, S154, T193, S203, and S287; eight potential protein kinase C phosphorylation sites at S57, T69, T95, S179, T269, S274, S275, and S284; and a potential signal peptide sequence from M1 to G27. SIGP-51 shares 26% identity with a human antigen precursor protein (GI 1814277); the protein kinase C phosphorylation sites at residues S57 and T69; and the casein kinase II phosphorylation site at residue T100. The fragment of SEQ ID NO:128 encoding the potential extracellular ligand binding domain from about nucleotide 88 to about nucleotide 148 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal,

and cardiovascular cDNA libraries. Approximately 48% of these libraries are associated with cancer, 29% are associated with immune response, and 20% with fetal disorders.

Nucleic acids encoding the SIGP-52 of the present invention were first identified in Incyte Clone 2452208 from the cardiovascular cDNA library (ENDANOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:129, was derived from Incyte Clones 2452280 (ENDANOT01), 1505094 (BRAITUT07), 1521239 (BLADTUT04), and 1309844 (COLNFET02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:52. SIGP-52 is 351 amino acids in length and has two potential N-glycosylation sites at N241 and N337; two potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at S201 and T318; six potential casein kinase II phosphorylation sites at S9, S136, T162, T252, S270, and S302; eight potential protein kinase C phosphorylation sites at T25, S34, T37, S64, S87, S112, S141, and S322; and one potential cell attachment sequence at R280GD. The fragment of SEQ ID NO:129 encoding the potential extracellular ligand binding domain from about nucleotide 97 to about nucleotide 157 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal, cardiovascular, and neural cDNA libraries. Approximately 33% of these libraries are associated with cancer, 33% are associated with immune response, and 26% with fetal disorders.

Nucleic acids encoding the SIGP-53 of the present invention were first identified in Incyte Clone 2457825 from the aortic endothelial cell cDNA library (ENDANOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:130, was derived from Incyte Clone 2457825 (ENDANOT01) and shotgun sequences SASA00641, SASA02817, SASA01973, SASA03121, SASA01350, and SASA00693.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:53. SIGP-53 is 662 amino acids in length and has three potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at S555, S578, and S652; ten potential casein kinase II phosphorylation sites at S67, T151, T215, S241, S470, S471, S482, S556, T589, and T618; one potential leucine zipper pattern from L572 to L593; four potential protein kinase C phosphorylation sites at T2, T21, S80, and T503;

and one potential LIM domain signature site from C402 to L436. SIGP-53 shares 10% identity with the C. elegans protein encoded by W04D2.1 (GI 1418625); and the casein kinase II phosphorylation site at residue S241. The fragment of SEQ ID NO:130 encoding the potential extracellular ligand binding domain from about nucleotide 88 to about nucleotide 148 is useful for hybridization. Northern analysis shows the expression of this sequence in hematopoietic, gastrointestinal, reproductive, and cardiovascular cDNA libraries. Approximately 43% of these libraries are associated with cancer, 35% are associated with immune response, and 22% with fetal disorders.

Nucleic acids encoding the SIGP-54 of the present invention were first identified in Incyte Clone 2470740 from the hematopoietic cDNA library (THP1NOT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:131, was derived from Incyte Clone 2470740 (THP1NOT03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:54. SIGP-54 is 115 amino acids in length and has one potential protein kinase C phosphorylation site at S85; and one potential insulin family signature site from C23 to C37. The fragment of SEQ ID NO:131 encoding the potential extracellular ligand binding domain from about nucleotide 151 to about nucleotide 211 is useful for hybridization. Northern analysis shows the expression of this sequence in neural and developmental cDNA libraries. Approximately 33% of these libraries are associated with cancer and 33% are associated with fetal disorders.

Nucleic acids encoding the SIGP-55 of the present invention were first identified in Incyte Clone 2479092 from the aortic endothelial cell cDNA library (SMCANOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:132, was derived from Incyte Clone 2479092 (SMCANOT01) and 1981954 (LUNGTUT03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:55. SIGP-55 is 157 amino acids in length and has one potential casein kinase II phosphorylation site at S31; one potential tyrosine kinase phosphorylation site at K150; and a potential signal peptide sequence from M1 to A26.

The fragment of SEQ ID NO:132 encoding the potential extracellular ligand binding

domain from about nucleotide 97 to about nucleotide 157 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal, hematopoietic, and urologic cDNA libraries. Approximately 47% of these libraries are associated with cancer and 29% with immune response.

5 Nucleic acids encoding the SIGP-56 of the present invention were first identified in Incyte Clone 2480544 from the aortic smooth muscle cell cDNA library (SMCANOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:133, was derived from Incyte Clones 2480544 (SMCANOT01), 2472409 (THP1NOT03), 1516031 (PANCTUT01), 855817 (NGANNOT01), 1865287  
10 (PROSNOT19), and 677835 (CRBLNOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:56. SIGP-56 is 197 amino acids in length and has one potential N glycosylation site at N38; one potential casein kinase II phosphorylation site at S123; two potential protein kinase C phosphorylation sites at T71 and S82; and a potential signal  
15 peptide sequence from M1 to A27. SIGP-56 shares 15% identity with a Phaseolus vulgaris protein involved in the stress response (GI 169345) and shows conservation of proline and tyrosine residues in the C-terminal region. The fragment of SEQ ID NO:133 from about nucleotide 125 to about nucleotide 160 is useful for hybridization. Northern analysis shows the expression of this sequence in neural, reproductive, and cardiovascular cDNA libraries.  
20 Approximately 49% of these libraries are associated with neoplastic disorders and 14% with immune response.

Nucleic acids encoding the SIGP-57 of the present invention were first identified in Incyte Clone 2518547 from the brain tumor cDNA library (BRAITUT21) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:134, was  
25 derived from Incyte Clones 2518547 (BRAITUT21), 1509622 (LUNGNOT14), 1562945 (SPLNNOT04), 1640136 (UTRSNOT06), and 1432014 (BEPINON01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:57. SIGP-57 is 245 amino acids in length and has one potential casein kinase II phosphorylation site at S27; and two potential protein kinase C  
30 phosphorylation sites at S5 and T229. SIGP-57 shares 36% identity with a human protein

that binds a regulatory element of the c-myc gene (GI 33969). In addition, the potential protein kinase C phosphorylation site at T229 is conserved as a potential protein kinase A phosphorylation site at S176 in the human protein. The fragment of SEQ ID NO:134 from about nucleotide 742 to about nucleotide 775 is useful for hybridization. Northern analysis shows the expression of this sequence in hematopoietic, reproductive, and neural cDNA libraries. Approximately 50% of these libraries are associated with neoplastic disorders and 28% with immune response.

Nucleic acids encoding the SIGP-58 of the present invention were first identified in Incyte Clone 2530650 from the gallbladder cDNA library (GBLANOT02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:135, was derived from Incyte Clones 2530650 (GBLANOT02), 2617724 (GBLANOT01), 3105644 (BRSTTUT15), 2903466 (DRGCNOT01), 1545010 (PROSTUT04), 2313837 (NGANNOT01), 1804413 (SINTNOT13), 3207379 (PENCNOT03), 2347051 (TESTTUT02), 2602493 (UTRSNOT10), 1259341 (MENITUT03), and 81943 (SYNORAB01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:58. SIGP-58 is 310 amino acids in length and has one potential N glycosylation site at N206; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at T97; five potential casein kinase II phosphorylation sites at S62, S156, S214, S222, and T274; five potential protein kinase C phosphorylation sites at T150, T167, T208, T265, and S273; one potential tyrosine kinase phosphorylation site at Y96; one thyroglobulin type-1 repeat signature from F109 to G143; and a potential signal peptide sequence from M1 to A21. SIGP-58 shares 18% identity with bovine thyroglobulin (GI 2204111) and 46% identity between F109 and G143, the thyroglobulin type-1 repeat signature. The fragment of SEQ ID NO:135 from about nucleotide 92 to about nucleotide 127 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive and cardiovascular cDNA libraries. Approximately 67% of these libraries are associated with neoplastic disorders and 19% with immune response.

Nucleic acids encoding the SIGP-59 of the present invention were first identified in Incyte Clone 2652271 from the thymus cDNA library (THYMNOT04) using a computer

search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:136, was derived from Incyte Clones 2652271 (THYMNOT04), 2742813 (BRSTTUT14), 763431 (BRAITUT02), 1272403 (TESTTUT02), 1240531 (LUNGNOT03), and 1318448 (BLADNOT04).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:59. SIGP-59 is 256 amino acids in length and has three potential N glycosylation sites at N76, N106, and N212; three potential casein kinase II phosphorylation sites at T46, S188, and T204; two potential protein kinase C phosphorylation sites at S130 and S221; two potential ribonuclease T2 family histidine active sites from W62 to P69 and from F110 to C121; and a potential signal peptide sequence from M1 to A24. SIGP-59 shares 24% identity with Solanum lycopersicum ribonuclease LE (GI 895855); 80% identity between W62 and P75, one of the two ribonuclease T2 family histidine active sites; and 92% identity between F110 and C121, the second of the two ribonuclease T2 family histidine active sites. The fragment of SEQ ID NO:136 from about nucleotide 462 to about nucleotide 494 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, hematopoietic, and gastrointestinal cDNA libraries. Approximately 53% of these libraries are associated with neoplastic disorders and 28% with immune response.

Nucleic acids encoding the SIGP-60 of the present invention were first identified in Incyte Clone 2746976 from the lung tumor cDNA library (LUNGTUT11) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:137, was derived from Incyte Clones 2746976 (LUNGTUT11), 488049 (HNT2AGT01), 1907738 (CONNTUT01), 782645 (MYOMNOT01), and 823864 (PROSNOT06).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:60. SIGP-60 is 160 amino acids in length and has one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at T31; four potential casein kinase II phosphorylation sites at S23, S47, S96, and S152; four potential protein kinase C phosphorylation sites at S23, T125, S126, and T149; and a clathrin adaptor complex small chain signature from I56 to F66. SIGP-60 shares 84% identity with mouse clathrin-associated protein 19 (GI 191983) and 91% identity with the clathrin adaptor complex small



chain signature between I56 and F66. In addition, all potential casein kinase II and protein kinase C phosphorylation sites are conserved between SIGP-60 and the mouse protein. The fragments of SEQ ID NO:137 from about nucleotide 144 to about nucleotide 170 and from about nucleotide 495 to about nucleotide 521 are useful for hybridization. Northern analysis shows the expression of this sequence in hematopoietic, cardiovascular, and reproductive cDNA libraries. Approximately 39% of these libraries are associated with neoplastic disorders and 39% with immune response.

Nucleic acids encoding the SIGP-61 of the present invention were first identified in Incyte Clone 2753496 from the THP-1 promonocyte cDNA library (THP1AZS08) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:138, was derived from Incyte Clones 2753496 (THP1AZS08), 2642512 (LUNGTUT08), 1367244 (SCORNON02), 474458 (MMLR1DT01), 1349777 (LATRTUT02), 1380831 (BRAITUT08), and 832934 (PROSTUT04).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:61. SIGP-61 is 341 amino acids in length and has one potential N glycosylation site at N66; four potential casein kinase II phosphorylation sites at T157, T207, S296, and S335; two potential protein kinase C phosphorylation sites at S159 and S296; and one potential tyrosine kinase phosphorylation site at Y184. SIGP-61 shares 17% identity with Schizosaccharomyces pombe BEM46, a protein involved in cell polarity (GI 987286) and the potential phosphorylation sites at T157 and S296. The fragment of SEQ ID NO:138 from about nucleotide 79 to about nucleotide 114 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal, and neural cDNA libraries. Approximately 52% of these libraries are associated with neoplastic disorders and 25% with immune response.

Nucleic acids encoding the SIGP-62 of the present invention were first identified in Incyte Clone 2781553 from the ovarian tumor cDNA library (OVRTUT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:139, was derived from Incyte Clones 2781553 (OVRTUT03), 1413079 (BRAINOT12), 894971 (BRSTNOT05), 2696043 (UTRSNOT12), 1267806 (BRAINOT09), 1961608 (BRSTNOT04), 1755817 (LIVRTUT01), 1793882 (PROSTUT05), 1251515

(LUNGFET03), 1560984 (SPLNNOT04), and 1872574 (LEUKNOT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:62. SIGP-62 is 430 amino acids in length and has one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S387; thirteen potential casein kinase II phosphorylation sites at S182, S214, S235, T248, S258, T266, T275, T294, S313, T356, S387, T404, and S413; six potential protein kinase C phosphorylation sites at T71, S168, S235, S306, T356, and S374; and a mitochondrial energy transfer protein signature from P114 to L122. Northern analysis shows the expression of this sequence in reproductive, neural, and hematopoietic cDNA libraries. Approximately 47% of these libraries are associated with neoplastic disorders and 19% with immune response.

Nucleic acids encoding the SIGP-63 of the present invention were first identified in Incyte Clone 2821925 from the adrenal tumor cDNA library (ADRETUT06) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:140, was derived from Incyte Clones 2821925 (ADRETUT06), 933799 (CERVNOT01), and 136467 (SYNORAB01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:63. SIGP-63 is 143 amino acids in length and has one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S109; three potential casein kinase II phosphorylation sites at S36, S80, and T84; five potential protein kinase C phosphorylation sites at T31, T55, T70, S109, and T122; and a potential signal peptide sequence from M1 to A21. Northern analysis shows the expression of this sequence in reproductive, musculoskeletal and cardiovascular cDNA libraries. Approximately 50% of these libraries are associated with neoplastic disorders and 27% with immune response.

Nucleic acids encoding the SIGP-64 of the present invention were first identified in Incyte Clone 2879068 from the uterine tumor cDNA library (UTRSTUT05) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:141, was derived from Incyte Clones 2879068 (UTRSTUT05), 2910155 (KIDNTUT15), 488673 (HNT2AGT01), 1285407 (COLNNOT16), 1415890 (BRAINOT12), 1352662 (LATRTUT02), 41046 (TBLYNOT01), and 2686554 (LUNGNOT23).

In one embodiment, the invention encompasses a polypeptide comprising the amino

acid sequence of SEQ ID NO:64. SIGP-64 is 301 amino acids in length and has two potential N glycosylation sites at N20 and N251; five potential casein kinase II phosphorylation sites at S8, S41, T125, T161, and T163; five potential protein kinase C phosphorylation sites at T40, S41, T59, T66, and S181; one potential tyrosine kinase phosphorylation site at Y176; one potential glycosaminoglycan attachment site at S253; and two putative RNP-1 RNA-binding signatures from R70 to F77 and from R155 to Y162. SIGP-64 shares 59% identity with human heterogeneous nuclear ribonucleoprotein D (GI 870749); 100% identity between R70 and F77, one of the two RNP-1 RNA-binding signatures; and 89% identity between R155 and Y162, the second of the two RNP-1 RNA-binding signatures. In addition, eight potential phosphorylation sites are conserved between SIGP-64 and the human ribonucleoprotein. The fragments of SEQ ID NO:141 from about nucleotide 207 to about nucleotide 248 and from about nucleotide 726 to about nucleotide 752 are useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, hematopoietic, and gastrointestinal cDNA libraries. Approximately 48% of these libraries are associated with neoplastic disorders and 24% with immune response.

Nucleic acids encoding the SIGP-65 of the present invention were first identified in Incyte Clone 2886757 from the small intestine cDNA library (SINJNOT02) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:142, was derived from Incyte Clones 2886757 (SINJNOT02), 2230747 (PROSNOT16), and 899432 (BRSTTUT03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:65. SIGP-65 is 233 amino acids in length and has two potential N-glycosylation sites at N82 and N196; one potential casein kinase II phosphorylation site at S170; and two potential protein kinase C phosphorylation sites at S102 and T134. SIGP-65 shares 22% identity with S. cerevisiae protein encoded by YOL135c (GI 1420026), and the potential casein kinase II phosphorylation site at S170 is conserved between the two proteins. The fragment of SEQ ID NO:142 from about nucleotide 99 to about nucleotide 137 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, cardiovascular, and gastrointestinal cDNA libraries. Approximately 59% of these libraries are associated with neoplastic disorders.

Nucleic acids encoding the SIGP-66 of the present invention were first identified in Incyte Clone 2964329 from the cervical spinal cord cDNA library (SCORNOT04) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:143, was derived from Incyte Clones 2964329, (SCORNOT04), 1274814 (TESTTUT02), 746049 (BRAITUT01), 1395667 (THYRNOT03), 1362944 (LUNGNOT12), and 2589 (HMC1NOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:66. SIGP-66 is 354 amino acids in length and has one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S346; two potential casein kinase II phosphorylation sites at S164 and T180; six potential protein kinase C phosphorylation sites at S43, S135, S150, S164, S172, and S201; and one potential tyrosine kinase phosphorylation site at Y182. SIGP-66 shares 12% identity with *S. cerevisiae* mitochondrial internal membrane carrier protein (GI 311667). In addition, one potential protein kinase C site is conserved between these molecules. The fragment of SEQ ID NO:143 from about nucleotide 416 to about nucleotide 442 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, hematopoietic/immune, gastrointestinal, and cardiovascular cDNA libraries. Approximately 46% of these libraries are associated with neoplastic disorders and 26% with immune response.

Nucleic acids encoding the SIGP-67 of the present invention were first identified in Incyte Clone 2965248 from the cervical spinal cord cDNA library (SCORNOT04) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:144, was derived from Incyte Clones 2965248 (SCORNOT04), 485746 (HNT2RAT01), 865684 (BRAITUT03), 1459157 (COLNFET02), 1597772 (BRAINOT14), 531430 (BRAINOT03), 725362 (SYNOOAT01), 1620429 (BRAITUT13), and 190305 (SYNORAB01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:67. SIGP-67 is 235 amino acids in length and has seven potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at S50, T80, T98, T126, S135, S136, and T194; three potential casein kinase II phosphorylation sites at

S60, T80, and S81; six potential protein kinase C phosphorylation sites at S114, T119, T137, S142, S146, and S174; and a strathmin 1 family signature from P75 to E84. SIGP-67 shares 44% identity with human strathmin homolog SCG10/neuron-specific growth-associated protein in Alzheimer's disease (GI 1478503), and 71% identity between M1 and A107. In addition, one potential cAMP- and cGMP-dependent protein kinase phosphorylation site, one potential casein kinase II phosphorylation site, the strathmin 1 family signature, and the hydrophobic transmembrane domains are conserved between these molecules. TM1 extends from about L15 to about F25; and TM2, from about G196 to about P212. The fragments of SEQ ID NO:144 from about nucleotide 158 to about nucleotide 196 and from about nucleotide 614 to about nucleotide 643 are useful for hybridization. Northern analysis shows the expression of this sequence in neural, reproductive, gastrointestinal, and hematopoietic/immune cDNA libraries. Approximately 50% of these libraries are associated with neoplastic disorders and 19% with immune response.

Nucleic acids encoding the SIGP-68 of the present invention were first identified in Incyte Clone 3000534 from the Th2 T lymphocyte cDNA library (TLYMNOT06) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:145, was derived from Incyte Clones 3000534 (TLYMNOT06), 1830964 (THP1AZT01), 1329136 (PANCNOT07), and 2910083 (KIDNTUT15).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:68. SIGP-68 is 221 amino acids in length and has two potential casein kinase II phosphorylation sites at T31 and T70; one potential glycosaminoglycan attachment site at S62; three potential protein kinase C phosphorylation sites at T111, T146, and T199; and an endoplasmic reticulum targeting sequence at H218DEL. SIGP-68 shares 61% identity with the human stroma cell-derived secretory factor-2 (GI 1741868). In addition, one potential protein kinase C phosphorylation site and the hydrophobic transmembrane domains are conserved between these molecules. TM1 extends from about A10 to about G27; and TM2, from about T31 to about L45. The cysteines at C38, C92, C100, and C149 are conserved between both molecules. The fragments of SEQ ID NO:145 from about nucleotide 89 to about nucleotide 118 and from

about nucleotide 608 to about nucleotide 643 are useful for hybridization. Northern analysis shows the expression of this sequence in hematopoietic/immune, reproductive, cardiovascular, and gastrointestinal cDNA libraries. Approximately 41 % of these libraries are associated with neoplastic disorders and 31 % with immune response.

5 Nucleic acids encoding the SIGP-69 of the present invention were first identified in Incyte Clone 3046870 from the coronary artery cDNA library (HEAANOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:146, was derived from Incyte Clones 3046870 (HEAANOT01), 2719210 (THYRNOT09), 581291 (SATPFI006), 1961256 (BRSTNOT04), 2226972  
10 (SEMVNOT01), 2023351 (CONNNOT01), 1379008 (LUNGNOT10), and 1943136 (HIPONOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:69. SIGP-69 is 483 amino acids in length and has one potential N-glycosylation site at N178; ten potential casein kinase II phosphorylation sites at S16, S49, T60, T67, T92, T121, T170, T187, T250, and S431; and nine potential  
15 protein kinase C phosphorylation sites at S113, T170, T187, T194, S210, T265, S284, T355, and S431. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal, cardiovascular, and neural cDNA libraries. Approximately 49% of these libraries are associated with neoplastic disorders and 24% with immune response.

20 Nucleic acids encoding the SIGP-70 of the present invention were first identified in Incyte Clone 3057669 from the pons cDNA library (PONSAZT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:147, was derived from Incyte Clones 3057669 (PONSAZT01), 548211 (BEPINOT01), 3702516 (PENCNOT07), 3581270 (293TF3T01), 495191 (HNT2NOT01), 2784427 (BRSTNOT13),  
25 1515961 (PANCTUT01), 3552333 (SYNONOT01), 2838668 (DRGLNOT01), 14600680 (COLNFET02), and 285677 (EOSIHET02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:70. SIGP-70 is 371 amino acids in length and has three potential N-glycosylation sites at N70, N125, and N362; eleven potential casein kinase II  
30 phosphorylation sites at T22, S66, S72, S73, S102, T160, T201, T215, T278, T285, and

S316; seven potential protein kinase C phosphorylation sites at S72, T79, S99, T127, S134, S257, and T299; and one protein kinase signature and profile from L188 to F200. Northern analysis shows the expression of this sequence in gastrointestinal, reproductive, and neural cDNA libraries. Approximately 54% of these libraries are associated with  
5 neoplastic disorders and 14% with immune response.

Nucleic acids encoding the SIGP-71 of the present invention were first identified in Incyte Clone 3088178 from the aorta cDNA library (HEAONOT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:148, was derived from Incyte Clones 3088178 (HEAONOT03), 589421 (UTRSNOT01), 2059958  
10 (OVARNOT03), 1550631 (PROSNOT06), and 1271480 (TESTTUT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:71. SIGP-71 is 402 amino acids in length and has two potential N glycosylation sites at N13 and N366; two potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at T50 and S51; five potential casein kinase II phosphorylation sites at T50, S51, S52, S56, and S246; one potential glycosaminoglycan attachment site at S247; eight potential protein kinase C phosphorylation sites at T45, T46, S224, S240, S259, T279, S338, and S376; one potential tyrosine kinase phosphorylation site at Y273; and one beta-transducin family Trp-Asp repeat signature from V243 to V257. SIGP-71 shares 22% identity with S. cerevisiae protein encoded by HRE594 (GI 498997; truncated sequence). In addition, one potential N-glycosylation site, and two potential  
15 casein kinase II phosphorylation sites are conserved between these molecules. The fragment of SEQ ID NO:148 from about nucleotide 725 to about nucleotide 766 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, cardiovascular, and hematopoietic/immune cDNA libraries. Approximately 51% of  
20 these libraries are associated with neoplastic disorders and 23% with immune response.

Nucleic acids encoding the SIGP-72 of the present invention were first identified in Incyte Clone 3094321 from the breast cDNA library (BRSTNOT19) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:149, was derived from Incyte Clones 3094321 (BRSTNOT19), 2517422H1 (BRAITUT21), 2101110  
30 (BRAITUT02), 1303603 (PLACNOT02), 2675275 (KIDNNOT19), 1988065

(LUNGAST01), 34101 (THP1NOB01), 1815156 (PROSNOT20), 602724 (BRSTTUT01), and 1485067 (CORPNOT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:72. SIGP-72 is 640 amino acids in length and has four potential N-glycosylation sites at N295, N513, N568, and N619; two potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at S239 and S507; sixteen potential casein kinase II phosphorylation sites at S42, T178, T220, S229, S239, T247, S289, S350, S372, S446, T463, S492, T580, S592, S604, and S625; nine potential protein kinase C phosphorylation sites at T150, T166, T174, S239, T328, S407, T451, S609, and S621; one potential tyrosine kinase phosphorylation site at Y265; and one cytochrome c family heme-binding site signature at C158YECHP. SIGP-72 shares 33% identity with an essential yeast ubiquitin-activating enzyme homolog (GI 793879). In addition, one potential N-glycosylation site, one potential casein kinase II phosphorylation site, and six potential protein kinase C phosphorylation sites are conserved between these molecules. The fragments of SEQ ID NO:149 from about nucleotide 382 to about nucleotide 423 and from about nucleotide 1087 to about nucleotide 1113 are useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, hematopoietic/immune, cardiovascular, and gastrointestinal cDNA libraries. Approximately 48% of these libraries are associated with neoplastic disorders and 24% with immune response.

Nucleic acids encoding the SIGP-73 of the present invention were first identified in Incyte Clone 3115936 from the lung cDNA library (LUNGTUT13) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:150, was derived from Incyte Clones 3115936 (LUNGTUT13) 2359411 (LUNGFET05), 2189762 (PROSNOT26), 1449756 (PLACNOT02), 541212 (LNODNOT02), 079364 (SYNORAB01), 864877 (BRAITUT03), 2697958 (UTRSNOT12), 1818830 (PROSNOT20), 1966765 (BRSTNOT04), 998279 (KIDNTUT01), 1961616 (BRSTNOT04), and 1431515 (BEPINON01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:73. SIGP-73 is 237 amino acids in length and has five potential casein kinase II phosphorylation sites at S43, S47, S72, S131, and T177; and



three potential protein kinase C phosphorylation sites at S39, S125, and T202. SIGP-73 shares 44% identity with t yeast Rer1p protein, which ensures correct localization of Sec12p integral membrane protein of the endoplasmic reticulum (GI 517174). In addition, the hydrophobic transmembrane domains are conserved among these molecules. TM1 extends from about A82 to about P126; and TM2, from about A166 to about M203. The fragment of SEQ ID NO:150 from about nucleotide 585 to about nucleotide 623 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, cardiovascular, gastrointestinal, and hematopoietic/ immune cDNA libraries. Approximately 48% of these libraries are associated with neoplastic disorders and 24% with immune response.

Nucleic acids encoding the SIGP-74 of the present invention were first identified in Incyte Clone 3116522 from the lung cDNA library (LUNGTUT13) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:151, was derived from Incyte Clones 3116522 (LUNGTUT13), 2523149 (BRAITUT21), 1513583 (PANCTUT01), 834017 (PROSNOT07), 1631796 (COLNNOT19), 1502736 (BRAITUT07), and 78850 (SYNORAB01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:74. SIGP-74 is 432 amino acids in length and has three potential casein kinase II phosphorylation sites at S144, S257, and S317; three potential protein kinase C phosphorylation sites at T68, S231, and T372; and one potential tyrosine kinase phosphorylation site at Y240. SIGP-74 shares 28% identity with the human UDP-galactose transporter isoform (GI 1669560). In addition, one potential protein kinase C phosphorylation site and the hydrophobic transmembrane domains are conserved between these molecules. TM4 extends from about Q108 to about G127; TM5, from about S152 to about L173; TM6, from about K205 to about K228; TM7, from about T242 to about S257; TM8, from about T268 to about S283; TM9, from about A294 to about T328; and TM10, from about A338 to about V409. The fragment of SEQ ID NO:151 from about nucleotide 710 to about nucleotide 736 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, gastrointestinal, cardiovascular, hematopoietic/immune, and urologic cDNA libraries. Approximately 54% of these

libraries are associated with neoplastic disorders and 25% with immune response.

Nucleic acids encoding the SIGP-75 of the present invention were first identified in Incyte Clone 3117184 from the lung cDNA library (LUNGTUT13) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:152, was derived from Incyte Clones 3117184 (LUNGTUT13), 2494724 (ADRETUT05), and 1922002 (BRSTTUT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:75. SIGP-75 is 252 amino acids in length and has one potential N-glycosylation site at N93; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at S179; one potential casein kinase II phosphorylation site at T189; and five potential protein kinase C phosphorylation sites at S95, S115, S123, T140, and T200. SIGP-75 shares 39% identity with C. elegans protein encoded by WO4D2.6 (GI 1418628). In addition, one potential N-glycosylation site, and three potential protein kinase C phosphorylation sites are conserved between the molecules. The fragment of SEQ ID NO:152 from about nucleotide 567 to about nucleotide 593 is useful for hybridization. Northern analysis shows the expression of this sequence in cardiovascular, gastrointestinal, hematopoietic/immune, and reproductive cDNA libraries. Approximately 50% of these libraries are associated with neoplastic disorders and 20% with immune response.

Nucleic acids encoding the SIGP-76 of the present invention were first identified in Incyte Clone 3125156 from the lymph node cDNA library (LNODNOT05) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:153, was derived from Incyte Clones 3125156 (LNODNOT05), 1417459 (BRAINOT12), 1567861 (UTRSNOT05), 154233 (THP1PLB02), 872652 (LUNGAST01), 2525803 (BRAITUT21), and 1209172 (BRSTNOT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:76. SIGP-76 is 523 amino acids in length and has one potential N glycosylation sites at N186; nine potential casein kinase II phosphorylation sites at S63, T85, S179, S188, T210, S231, T269, T295, and S474; one potential glycosaminoglycan attachment site at S335; ten potential protein kinase C phosphorylation sites at T9, S159, S172, S179, T246, S263, S283, S416, S447, and S498; two potential tyrosine kinase

phosphorylation sites at Y106 and Y170; and one tyrosine specific protein phosphatase active site at V331. SIGP-76 shares 21% identity with human T-cell protein tyrosine phosphatase (GI 804750), the N186 glycosylation site, the phosphorylation sites at S179, S188, T210, T246, S263, T295, S416, and Y170; and 50% identity between P324 and F344, the region of the tyrosine specific protein phosphatase active site. The fragments of SEQ ID NO:153 from about nucleotide 64 to about nucleotide 183 and from about nucleotide 1087 to about nucleotide 1119 are useful for hybridization. Northern analysis shows the expression of this sequence in neural, reproductive, and gastrointestinal cDNA libraries. Approximately 55% of these libraries are associated with neoplastic disorders and 22% with immune response.

Nucleic acids encoding the SIGP-77 of the present invention were first identified in Incyte Clone 3129120 from the lung tumor cDNA library (LUNGTUT12) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:154, was derived from Incyte Clones 3129120 (LUNGTUT12), 3744590 (THYMNOT08), 1512939 (PANCTUT01), 3220539 (COLNNON03), 1435889 (PANCNOT08), 1452745 (PENITUT01), 874548 (LUNGAST01), 1524326 (UCMCL5T01), and 811239 (LUNGNOT04).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:77. SIGP-77 is 621 amino acids in length and has two potential N glycosylation sites at N203 and N517; one potential protein kinase A or G phosphorylation site at S84; five potential casein kinase II phosphorylation sites at T45, T185, T233, T278, and S573; seven potential protein kinase C phosphorylation sites at T45, T95, S109, S299, T318, S324, and T482; and one potential leucine zipper motif from L332 to L353. SIGP-77 shares 27% identity and the phosphorylation site at T318 with S. cerevisiae membrane protein important for endocytosis (GI 1256890). The fragments of SEQ ID NO:154 from about nucleotide 64 to about nucleotide 183 and from about nucleotide 1087 to about nucleotide 1119 are useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, gastrointestinal, and cardiovascular cDNA libraries. Approximately 53% of these libraries are associated with neoplastic disorders and 17% with immune response.

The invention also encompasses SIGP variants. A preferred SIGP variant is one which

has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the SIGP amino acid sequence, and which contains at least one functional or structural characteristic of SIGP.

The invention also encompasses polynucleotides which encode SIGP. Accordingly, any  
5 nucleic acid sequence which encodes the amino acid sequence of SIGP can be used to produce recombinant molecules which express SIGP. In a particular embodiment, the invention encompasses a polynucleotide consisting of a nucleic acid sequence selected from the group consisting of SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID  
10 NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID  
15 NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID  
20 NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the  
25 genetic code, a multitude of polynucleotide sequences encoding SIGP, some bearing minimal homology to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic  
30 code as applied to the polynucleotide sequence of naturally occurring SIGP, and all such

variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode SIGP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring SIGP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding SIGP or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding SIGP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode SIGP and SIGP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding SIGP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139,

SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144,  
 SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149,  
 SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154,  
 under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987)  
 5 Methods Enzymol. 152:399-407; and Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.)

Methods for DNA sequencing are well known and generally available in the art and may  
 be used to practice any of the embodiments of the invention. The methods may employ such  
 enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp.,  
 Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham,  
 10 Chicago, IL), or combinations of polymerases and proofreading exonucleases such as those  
 found in the ELONGASE Amplification System (GIBCO/BRL, Gaithersburg, MD).  
 Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200  
 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and  
 the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

15 The nucleic acid sequences encoding SIGP may be extended utilizing a partial nucleotide  
 sequence and employing various methods known in the art to detect upstream sequences,  
 such as promoters and regulatory elements. For example, one method which may be  
 employed, restriction-site PCR, uses universal primers to retrieve unknown sequence adjacent  
 to a known locus. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) In  
 20 particular, genomic DNA is first amplified in the presence of a primer complementary to a  
 linker sequence within the vector and a primer specific to the region predicted to encode the  
 gene. The amplified sequences are then subjected to a second round of PCR with the same  
 linker primer and another specific primer internal to the first one. Products of each round of  
 PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse  
 25 transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers  
 based on a known region. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.)  
 The primers may be designed using commercially available software such as OLIGO 4.06  
 Primer Analysis software (National Biosciences Inc., Plymouth, MN) or another appropriate  
 30 program to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or

more, and to anneal to the target sequence at temperatures of about 68°C to 72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

5 Another method which may be used is capture PCR, which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before  
10 performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060.) Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

15 When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable in that they will include more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into  
20 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated,  
25 and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., Genotyper™ and Sequence Navigator™, Perkin Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which  
30 might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode SIGP may be used in recombinant DNA molecules to direct expression of SIGP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express SIGP.

As will be understood by those of skill in the art, it may be advantageous to produce SIGP-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter SIGP-encoding sequences for a variety of reasons including, but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding SIGP may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of SIGP activity, it may be useful to encode a chimeric SIGP protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the SIGP encoding sequence and the heterologous protein sequence, so that SIGP may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding SIGP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser.



225-232.) Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of SIGP, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A Peptide Synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1983) Proteins, Structures and Molecular Properties, WH Freeman and Co., New York, NY.) Additionally, the amino acid sequence of SIGP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active SIGP, the nucleotide sequences encoding SIGP or derivatives thereof may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding SIGP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, ch. 4, 8, and 16-17; and Ausubel, F.M. et al. (1995, and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding SIGP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV))

or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions, e.g., enhancers, promoters, and 5' and 3' untranslated regions, of the vector and polynucleotide sequences encoding SIGP which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters, e.g., hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, La Jolla, CA) or pSport1™ plasmid (GIBCO/BRL), may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding SIGP, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for SIGP. For example, when large quantities of SIGP are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding SIGP may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of  $\beta$ -galactosidase so that a hybrid protein is produced, and pIN vectors. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) pGEX vectors (Pharmacia Biotech, Uppsala, Sweden) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or

factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used. (See, e.g., Ausubel, supra; and Grant et al. (1987) Methods Enzymol. 153:516-544.)

In cases where plant expression vectors are used, the expression of sequences encoding SIGP may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews. (See, e.g., Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.)

An insect system may also be used to express SIGP. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding SIGP may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of sequences encoding SIGP will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which SIGP may be expressed. (See, e.g., Engelhard, E.K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227.)

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding SIGP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing SIGP in infected

host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding SIGP. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding SIGP and its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding, and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long term, high yield production of recombinant proteins, stable expression is

preferred. For example, cell lines capable of stably expressing SIGP can be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector.

Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase genes and adenine phosphoribosyltransferase genes, which can be employed in *tk* or *ap<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; and Lowy, I. et al. (1980) Cell 22:817-823) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *npt* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g., *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Recently, the use of visible markers has gained popularity with such markers as anthocyanins,  $\beta$  glucuronidase and its substrate GUS, luciferase and its substrate luciferin. Green fluorescent proteins (GFP) (Clontech, Palo Alto, CA) are also used (See, e.g., Chalfie, M. et al. (1994) Science 263:802-805.) These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed.

For example, if the sequence encoding SIGP is inserted within a marker gene sequence, transformed cells containing sequences encoding SIGP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding SIGP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding SIGP and express SIGP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

The presence of polynucleotide sequences encoding SIGP can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding SIGP. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding SIGP to detect transformants containing DNA or RNA encoding SIGP.

A variety of protocols for detecting and measuring the expression of SIGP, using either polyclonal or monoclonal antibodies specific for the protein, are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on SIGP is preferred, but a competitive binding assay may be employed. These and other assays are well described in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Section IV; and Maddox, D.E. et al. (1983) J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding SIGP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding SIGP, or any fragments thereof,

may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, MI), Promega (Madison, WI), and U.S. Biochemical Corp. (Cleveland, OH). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding SIGP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode SIGP may be designed to contain signal sequences which direct secretion of SIGP through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding SIGP to nucleotide sequences encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences, such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA), between the purification domain and the SIGP encoding sequence may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing SIGP and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on immobilized metal ion affinity chromatography. (IMAC) (See, e.g., Porath, J. et al. (1992) Prot. Exp. Purif. 3: 263-281.) The enterokinase cleavage site provides a means for purifying SIGP from the fusion protein. (See, e.g., Kroll, D.J. et al. (1993) DNA Cell Biol. 12:441-453.)

Fragments of SIGP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, T.E. (1984) Protein: Structures and Molecular Properties, pp. 55-60, W.H. Freeman and Co., New York, NY.) Protein synthesis may be performed by manual techniques or by automation.

5 Automated synthesis may be achieved, for example, using the Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of SIGP may be synthesized separately and then combined to produce the full length molecule.

# **THERAPEUTICS**

10 The expression of the human signal peptide-containing proteins of the invention (SIGP) is closely associated with cell proliferation. Therefore, in cancers or immune response where SIGP is an activator, transcription factor, or enhancer, and is promoting cell proliferation, it is desirable to decrease the expression of SIGP. In conditions where SIGP is an inhibitor or suppressor and is controlling or decreasing cell proliferation, it is desirable to provide the protein or to increase the expression of SIGP.

15 In one embodiment, where SIGP is an inhibitor, SIGP or a fragment or derivative thereof may be administered to a subject to treat or prevent a cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma. Such cancers include, but are not limited to, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

20 In another embodiment, a pharmaceutical composition comprising purified SIGP may be used to treat or prevent a cancer including, but not limited to, those listed above.

25 In another embodiment, an agonist which is specific for SIGP may be administered to a subject to treat or prevent a cancer including, but not limited to, those cancers listed above.

In another further embodiment, a vector capable of expressing SIGP, or a fragment or a derivative thereof, may be administered to a subject to treat or prevent a cancer including, but not limited to, those cancers listed above.

30 In a further embodiment where SIGP is promoting cell proliferation, antagonists which



decrease the expression or activity of SIGP may be administered to a subject to treat or prevent a cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma. Such cancers include, but are not limited to, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, antibodies which specifically bind SIGP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express SIGP.

In another embodiment, a vector expressing the complement of the polynucleotide encoding SIGP may be administered to a subject to treat or prevent a cancer including, but not limited to, those cancers listed above.

In yet another embodiment where SIGP is promoting leukocyte activity or proliferation, antagonists which decrease the activity of SIGP may be administered to a subject to treat or prevent an immune response. Such responses include, but are not limited to, disorders such as AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and trauma. In one aspect, antibodies which specifically bind SIGP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express SIGP.

In another embodiment, a vector expressing the complement of the polynucleotide encoding SIGP may be administered to a subject to treat or prevent an immune response including, but not limited to, those listed above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists,

complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of SIGP may be produced using methods which are generally known in the art. In particular, purified SIGP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind SIGP. Antibodies to SIGP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with SIGP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to SIGP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of SIGP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to SIGP may be prepared using any technique which provides for

the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci.* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce SIGP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) *Proc. Natl. Acad. Sci.* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci.* 86: 3833-3837; and Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for SIGP may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex

formation between SIGP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering SIGP epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

In another embodiment of the invention, the polynucleotides encoding SIGP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding SIGP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding SIGP. Thus, complementary molecules or fragments may be used to modulate SIGP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding SIGP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequences complementary to the polynucleotides of the gene encoding SIGP. (See, e.g., Sambrook, supra; and Ausubel, supra.)

Genes encoding SIGP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding SIGP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding SIGP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology.

Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, pp. 163-177.) A  
5 complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific  
10 hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding SIGP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified  
15 by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility  
20 to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by  
25 in vitro and in vivo transcription of DNA sequences encoding SIGP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life.

30 Possible modifications include, but are not limited to, the addition of flanking sequences at

the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of SIGP, antibodies to SIGP, and mimetics, agonists, antagonists, or inhibitors of SIGP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which

facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

5        Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

10        Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants;  
15        cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

20        Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

25        Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable  
30        liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium

5 carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or

10 agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a

15 manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and

20 succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

25 After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of SIGP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended

30 purpose. The determination of an effective dose is well within the capability of those skilled



in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example SIGP or fragments thereof, antibodies of SIGP, and agonists, antagonists or inhibitors of SIGP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, and it can be expressed as the ED<sub>50</sub>/LD<sub>50</sub> ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu$ g to 100,000  $\mu$ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in

the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## 5    **DIAGNOSTICS**

In another embodiment, antibodies which specifically bind SIGP may be used for the diagnosis of disorders characterized by expression of SIGP, or in assays to monitor patients being treated with SIGP or agonists, antagonists, or inhibitors of SIGP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics.

10    Diagnostic assays for SIGP include methods which utilize the antibody and a label to detect SIGP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

15    A variety of protocols for measuring SIGP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of SIGP expression. Normal or standard values for SIGP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to SIGP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means.  
20    Quantities of SIGP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding SIGP may be  
25    used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of SIGP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of SIGP, and to monitor regulation of  
30    SIGP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding SIGP or closely related molecules may be used to identify nucleic acid sequences which encode SIGP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding SIGP, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the SIGP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154, or from genomic sequences including promoters, enhancers, and introns of the SIGP gene.

Means for producing specific hybridization probes for DNAs encoding SIGP include the cloning of polynucleotide sequences encoding SIGP or SIGP derivatives into vectors for the

production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding SIGP may be used for the diagnosis of a disorder associated with either increased or decreased expression of SIGP. Examples of such a disorder include, but are not limited to, cancers such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and cancers of the adrenal gland, bladder, bone, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, bone marrow, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; neuronal disorders such as akathisia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, schizophrenia, and Tourette's disorder; and immunological disorders such as AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and thyroiditis. The polynucleotide sequences encoding SIGP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered SIGP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding SIGP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The

nucleotide sequences encoding SIGP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding SIGP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of SIGP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding SIGP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding SIGP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding SIGP, or a fragment of a polynucleotide complementary to the polynucleotide encoding SIGP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of SIGP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; and Duplaa, C. et al. (1993) Anal. Biochem. 229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

In one embodiment, the microarray is prepared and used according to methods known in the art. (See, e.g., Chee et al. (1995) PCT application WO95/11995; Lockhart, D. J. et al. (1996) Nat. Biotech. 14:1675-1680; and Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619.)

The microarray is preferably composed of a large number of unique single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs. The oligonucleotides are preferably about 6 to 60 nucleotides in length, more preferably about 15 to 30 nucleotides in length, and most preferably about 20 to 25 nucleotides in length. It may be preferable to use oligonucleotides which are about 7 to 10

nucleotides in length. The microarray may contain oligonucleotides which cover the known 5' or 3' sequence, sequential oligonucleotides which cover the full length sequence, or unique oligonucleotides selected from particular areas along the length of the sequence.

Polynucleotides used in the microarray may be oligonucleotides specific to a gene or genes of interest. Oligonucleotides can also be specific to one or more unidentified cDNAs associated with a particular cell type or tissue type. It may be appropriate to use pairs of oligonucleotides on a microarray. The first oligonucleotide in each pair differs from the second oligonucleotide by one nucleotide. This nucleotide is preferably located in the center of the sequence. The second oligonucleotide serves as a control. The number of oligonucleotide pairs may range from about 2 to 1,000,000.

In order to produce oligonucleotides for use on a microarray, the gene of interest is examined using a computer algorithm which starts at the 5' end, or, more preferably, at the 3' end of the nucleotide sequence. The algorithm identifies oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack secondary structure that may interfere with hybridization. In one aspect, the oligomers may be synthesized on a substrate using a light-directed chemical process. (See, e.g., Chee et al., supra.) The substrate may be any suitable solid support, e.g., paper, nylon, any other type of membrane, or a filter, chip, or glass slide.

In another aspect, the oligonucleotides may be synthesized on the surface of the substrate using a chemical coupling procedure and an ink jet application apparatus. (See, e.g., Baldeschweiler et al. (1995) PCT application WO95/251116.) An array analogous to a dot or slot blot (HYBRIDOT® apparatus, GIBCO/BRL) may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system or thermal, UV, mechanical, or chemical bonding procedures. An array may also be produced by hand or by using available devices, materials, and machines, e.g. Brinkmann® multichannel pipettors or robotic instruments. The array may contain from 2 to 1,000,000 or any other feasible number of oligonucleotides.

In order to conduct sample analysis using the microarrays, polynucleotides are extracted from a sample. The sample may be obtained from any bodily fluid, e.g., blood, urine, saliva, phlegm, gastric juices, cultured cells, biopsies, or other tissue preparations. To produce

probes, the polynucleotides extracted from the sample are used to produce nucleic acid sequences complementary to the nucleic acids on the microarray. If the microarray contains cDNAs, antisense RNAs (aRNAs) are appropriate probes. Therefore, in one aspect, mRNA is reverse-transcribed to cDNA. The cDNA, in the presence of fluorescent label, is used to produce fragment or oligonucleotide aRNA probes. The fluorescently labeled probes are incubated with the microarray so that the probes hybridize to the microarray oligonucleotides. Nucleic acid sequences used as probes can include polynucleotides, fragments, and complementary or antisense sequences produced using restriction enzymes, PCR, or other methods known in the art.

Hybridization conditions can be adjusted so that hybridization occurs with varying degrees of complementarity. A scanner can be used to determine the levels and patterns of fluorescence after removal of any nonhybridized probes. The degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray can be assessed through analysis of the scanned images. A detection system may be used to measure the absence, presence, or level of hybridization for any of the sequences. (See, e.g., Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155.)

In another embodiment of the invention, nucleic acid sequences encoding SIGP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R.A. (ed.) Molecular Biology and Biotechnology, VCH Publishers New York, NY, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding SIGP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA



associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques.

Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., AT to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, SIGP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between SIGP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with SIGP, or fragments thereof, and washed. Bound SIGP is then detected by methods well known in the art. Purified SIGP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which

neutralizing antibodies capable of binding SIGP specifically compete with a test compound for binding SIGP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with SIGP.

In additional embodiments, the nucleotide sequences which encode SIGP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

## EXAMPLES

For purposes of example, the preparation and sequencing of the SPLNNOT04 cDNA library, from which Incyte Clones 1534876 and 1559131 were isolated, is described. Preparation and sequencing of cDNAs in libraries in the LIFESEQ™ database have varied over time, and the gradual changes involved use of kits, plasmids, and machinery available at the particular time the library was made and analyzed.

### I. SPLNNOT04 cDNA Library Construction

The SPLNNOT04 cDNA library was constructed from microscopically normal spleen tissue obtained from a 2-year-old Hispanic male who died of cerebral anoxia. The patient's serologies and past medical history were negative.

The frozen tissue was homogenized and lysed using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments, Westbury, NJ) in guanidinium isothiocyanate solution. The lysate was centrifuged over a 5.7 M CsCl cushion using an Beckman SW28 rotor in a Beckman L8-70M Ultracentrifuge (Beckman Instruments) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted with acid phenol pH 4.0, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNase-free water and DNase treated at 37°C. The RNA extraction and precipitation were repeated as before. The mRNA was then isolated using the Qiagen Oligotex kit (QIAGEN Inc., Chatsworth, CA) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SuperScript plasmid system (Cat. #18248-013, GIBCO-BRL, Gaithersburg, MD). cDNA synthesis was initiated with a NotI-oligo d(T) primer. Double-stranded cDNA was blunted, ligated to EcoRI adaptors, digested with NotI, fractionated on a Sepharose CL4B column (Cat. #275105-01, Pharmacia), and those cDNAs exceeding 400 bp were ligated into the NotI and EcoRI sites of the pINCY 1 vector (Incyte). The plasmid pINCY 1 was subsequently transformed into DH5 $\alpha$ <sup>TM</sup> competent cells (Cat. #18258-012, GIBCO-BRL).

## II Isolation and Sequencing of cDNA Clones

Plasmid cDNA was released from the cells and purified using the REAL Prep 96 plasmid kit (Catalog #26173, QIAGEN). The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, GIBCO-BRL) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

cDNAs were sequenced according to the method of Sanger et al. (1975, J. Mol. Biol. 94:441f), using the Perkin Elmer Catalyst 800 or a Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown, MA) and Applied Biosystems 377 DNA Sequencing Systems or the Perkin Elmer 373 DNA Sequencing System and the reading frame was determined.

## III. Homology Searching of cDNA Clones and Their Deduced Proteins

The nucleotide sequences and/or amino acid sequences of the Sequence Listing were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II databases. These databases, which contain previously identified and annotated sequences, were searched for regions of homology using BLAST (Basic Local Alignment Search Tool). (See, e.g., Altschul, S.F. (1993) J. Mol. Evol 36:290-300; and Altschul et al. (1990) J. Mol. Biol. 215:403-410.)

BLAST produced alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms could have been used when dealing with primary sequence patterns and secondary structure gap penalties. (See, e.g., Smith, T. et al. (1992) Protein Engineering 5:35-51.) The sequences disclosed in this application have lengths of at least 49 nucleotides and have no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at  $10^{-25}$  for nucleotides and  $10^{-8}$  for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate (pri), rodent (rod), and other mammalian sequences (mam), and deduced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp), for homology.

#### IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; and Ausubel, F.M. et al. *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

100

110

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact.

Homologous molecules are usually identified by selecting those which show product scores  
5 between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding SIGP occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in  
10 the cDNA library.

## V. Extension of SIGP Encoding Polynucleotides

The nucleic acid sequence of one of the polynucleotides of the present invention was used to design oligonucleotide primers for extending a partial nucleotide sequence to full  
15 length. One primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 (National Biosciences, Plymouth, MN), or  
20 another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO/BRL) were used to extend the sequence. If  
25 more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA), beginning with  
30 40 pmol of each primer and the recommended concentrations of all other components of the

kit, with the following parameters:

Step 1	94° C for 1 min (initial denaturation)
Step 2	65° C for 1 min
Step 3	68° C for 6 min
Step 4	94° C for 15 sec
Step 5	65° C for 1 min
Step 6	68° C for 7 min
Step 7	Repeat steps 4 through 6 for an additional 15 cycles
Step 8	94° C for 15 sec
Step 9	65° C for 1 min
Step 10	68° C for 7:15 min
Step 11	Repeat steps 8 through 10 for an additional 12 cycles
Step 12	72° C for 8 min
Step 13	4° C (and holding)

A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQuick™ (QIAGEN Inc., Chatsworth, CA), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13  $\mu$ l of ligation buffer, 1  $\mu$ l T4-DNA ligase (15 units) and 1  $\mu$ l T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent *E. coli* cells (in 40  $\mu$ l of appropriate media) were transformed with 3  $\mu$ l of ligation mixture and cultured in 80  $\mu$ l of SOC medium. (See, e.g., Sambrook, *supra*, Appendix A, p. 2.) After incubation for one hour at 37° C, the *E. coli* mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, *supra*, Appendix A, p. 1) containing 2x Carb. The following day, several colonies were randomly picked from each plate and cultured in 150  $\mu$ l of liquid LB/2x Carb medium placed in an individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5  $\mu$ l of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5  $\mu$ l from each sample was transferred into a PCR array.

For PCR amplification, 18  $\mu$ l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using

the following conditions:

- |        |  |
|--------|--|
| Step 1 | 94° C for 60 sec                                     |
| Step 2 | 94° C for 20 sec                                     |
| Step 3 | 55° C for 30 sec                                     |
| Step 4 | 72° C for 90 sec                                     |
| Step 5 | Repeat steps 2 through 4 for an additional 29 cycles |
| Step 6 | 72° C for 180 sec                                    |
| Step 7 | 4° C (and holding)                                   |

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequence of one of the nucleotide sequences of the present invention were used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

## VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from one of the nucleotide sequences of the present invention are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham, Chicago, IL), and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified using a Sephadex G-25 superfine resin column (Pharmacia & Upjohn, Kalamazoo, MI). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN, Boston, MA).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x

saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester, NY) is exposed to the blots to film for several hours, hybridization patterns are compared visually.

## 5 VII. Microarrays

To produce oligonucleotides for a microarray, one of the nucleotide sequences of the present invention is examined using a computer algorithm which starts at the 3' end of the nucleotide sequence. For each, the algorithm identifies oligomers of defined length that are unique to the nucleic acid sequence, have a GC content within a range suitable for  
 10 hybridization, and lack secondary structure that would interfere with hybridization. The algorithm identifies approximately 20 oligonucleotides corresponding to each nucleic acid sequence. For each sequence-specific oligonucleotide, a pair of oligonucleotides is synthesized in which the first oligonucleotide differs from the second oligonucleotide by one nucleotide in the center of the sequence. The oligonucleotide pairs can be arranged on a substrate, e.g. a silicon chip, using a light-directed chemical process. (See, e.g., Chee, supra.)  
 15

In the alternative, a chemical coupling procedure and an ink jet device can be used to synthesize oligomers on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link fragments or oligonucleotides to the surface of a substrate using or thermal, UV, mechanical, or chemical  
 20 bonding procedures, or a vacuum system. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray may be assessed through analysis of the  
 25 scanned images.

## VIII. Complementary Polynucleotides

Sequences complementary to the SIGP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring SIGP. Although use of  
 30 oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same



procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using Oligo 4.06 software and the coding sequence of SIGP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the SIGP-encoding transcript.

## IX. Expression of SIGP

Expression of SIGP is accomplished by subcloning the cDNA into an appropriate vector and transforming the vector into host cells. This vector contains an appropriate promoter, e.g.,  $\beta$ -galactosidase upstream of the cloning site, operably associated with the cDNA of interest. (See, e.g., Sambrook, *supra*, pp. 404-433; and Rosenberg, M. et al. (1983) *Methods Enzymol.* 101:123-138.)

Induction of an isolated, transformed bacterial strain with isopropyl beta-D-thiogalactopyranoside (IPTG) using standard methods produces a fusion protein which consists of the first 8 residues of  $\beta$ -galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of SIGP into bacterial growth media which can be used directly in the following assay for activity.

## X. Production of SIGP Specific Antibodies

SIGP substantially purified using PAGE electrophoresis (see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The SIGP amino acid sequence is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel et al. *supra*, ch. 11.)

Typically, the oligopeptides are 15 residues in length, and are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and coupled to

KLH (Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel et al. supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radioiodinated goat anti-rabbit IgG.

## **XI. Purification of Naturally Occurring SIGP Using Specific Antibodies**

Naturally occurring or recombinant SIGP is substantially purified by immunoaffinity chromatography using antibodies specific for SIGP. An immunoaffinity column is constructed by covalently coupling anti-SIGP antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing SIGP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of SIGP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/SIGP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and SIGP is collected.

## **XII. Identification of Molecules Which Interact with SIGP**

SIGP, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled SIGP, washed, and any wells with labeled SIGP complex are assayed. Data obtained using different concentrations of SIGP are used to calculate values for the number, affinity, and association of SIGP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be

unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Lal, Preeti  
Hillman, Jennifer L.  
Corley, Neil C.  
Guegler, Karl J.  
Baugh, Mariah  
Sather, Susan  
Shah, Purvi
- (ii) TITLE OF INVENTION: HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS
- (iii) NUMBER OF SEQUENCES:154
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
  - (B) STREET: 3174 PORTER DRIVE
  - (C) CITY: PALO ALTO
  - (D) STATE: CALIFORNIA
  - (E) COUNTRY: USA
  - (F) ZIP: 94304
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: TO BE ASSIGNED
  - (B) FILING DATE: HERewith
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: BILLINGS, LUCY J.
  - (B) REGISTRATION NUMBER: 36,749
  - (C) REFERENCE/DOCKET NUMBER:PF-0459 US
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (650) 855-0555
  - (B) TELEFAX: (650) 845-4166

(2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 348 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:

PF-0459 US

(A) LIBRARY: HEARNOT01  
(B) CLONE: 305841

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1 :

Met	Ala	Ala	Thr	Leu	Gly	Pro	Leu	Gly	Ser	Trp	Gln	Gln	Trp	Arg	
				5					10					15	
Arg	Cys	Leu	Ser	Ala	Arg	Asp	Gly	Ser	Arg	Met	Leu	Leu	Leu	Leu	
				20					25					30	
Leu	Leu	Leu	Gly	Ser	Gly	Gln	Gly	Pro	Gln	Gln	Val	Gly	Ala	Gly	
				35					40					45	
Gln	Thr	Phe	Glu	Tyr	Leu	Lys	Arg	Glu	His	Ser	Leu	Ser	Lys	Pro	
				50					55					60	
Tyr	Gln	Gly	Val	Gly	Thr	Gly	Ser	Ser	Ser	Leu	Trp	Asn	Leu	Met	
				65					70					75	
Gly	Asn	Ala	Met	Val	Met	Thr	Gln	Tyr	Ile	Arg	Leu	Thr	Pro	Asp	
				80					85					90	
Met	Gln	Ser	Lys	Gln	Gly	Ala	Leu	Trp	Asn	Arg	Val	Pro	Cys	Phe	
				95					100					105	
Leu	Arg	Asp	Trp	Glu	Leu	Gln	Val	His	Phe	Lys	Ile	His	Gly	Gln	
				110					115					120	
Gly	Lys	Lys	Asn	Leu	His	Gly	Asp	Gly	Leu	Ala	Ile	Trp	Tyr	Thr	
				125					130					135	
Lys	Asp	Arg	Met	Gln	Pro	Gly	Pro	Val	Phe	Gly	Asn	Met	Asp	Lys	
				140					145					150	
Phe	Val	Gly	Leu	Gly	Val	Phe	Val	Asp	Thr	Tyr	Pro	Asn	Glu	Glu	
				155					160					165	
Lys	Gln	Gln	Glu	Arg	Val	Phe	Pro	Tyr	Ile	Ser	Ala	Met	Val	Asn	
				170					175					180	
Asn	Gly	Ser	Leu	Ser	Tyr	Asp	His	Glu	Arg	Asp	Gly	Arg	Pro	Thr	
				185					190					195	
Glu	Leu	Gly	Gly	Cys	Thr	Ala	Ile	Val	Arg	Asn	Leu	His	Tyr	Asp	
				200					205					210	
Thr	Phe	Leu	Val	Ile	Arg	Tyr	Val	Lys	Arg	His	Leu	Thr	Ile	Met	
				215					220					225	
Met	Asp	Ile	Asp	Gly	Lys	His	Glu	Trp	Arg	Asp	Cys	Ile	Glu	Val	
				230					235					240	
Pro	Gly	Val	Arg	Leu	Pro	Arg	Gly	Tyr	Tyr	Phe	Gly	Thr	Ser	Ser	
				245					250					255	
Ile	Thr	Gly	Asp	Leu	Ser	Asp	Asn	His	Asp	Val	Ile	Ser	Leu	Lys	
				260					265					270	
Leu	Phe	Glu	Leu	Thr	Val	Glu	Arg	Thr	Pro	Glu	Glu	Glu	Lys	Leu	
				275					280					285	
His	Arg	Asp	Val	Phe	Leu	Pro	Ser	Val	Asp	Asn	Met	Lys	Leu	Pro	
				290					295					300	
Glu	Met	Thr	Ala	Pro	Leu	Pro	Pro	Leu	Ser	Gly	Leu	Ala	Leu	Phe	
				305					310					315	
Leu	Ile	Val	Phe	Phe	Ser	Leu	Val	Phe	Ser	Val	Phe	Ala	Ile	Val	
				320					325					330	
Ile	Gly	Ile	Ile	Leu	Tyr	Asn	Lys	Trp	Gln	Glu	Gln	Ser	Arg	Lys	
				335					340					345	
Arg	Phe	Tyr													

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 194 amino acids  
(B) TYPE: amino acid

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: EOSIHET02  
(B) CLONE: 322866

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2 :

Met	Gly	Met	Ser	Ser	Leu	Lys	Leu	Leu	Lys	Tyr	Val	Leu	Phe	Phe	
				5					10					15	
Phe	Asn	Leu	Leu	Phe	Trp	Ile	Cys	Gly	Cys	Cys	Ile	Leu	Gly	Phe	
				20					25					30	
Gly	Ile	Tyr	Leu	Leu	Ile	His	Asn	Asn	Phe	Gly	Val	Leu	Phe	His	
				35					40					45	
Asn	Leu	Pro	Ser	Leu	Thr	Leu	Gly	Asn	Val	Phe	Val	Ile	Val	Gly	
				50					55					60	
Ser	Ile	Ile	Met	Val	Val	Ala	Phe	Leu	Gly	Cys	Met	Gly	Ser	Ile	
				65					70					75	
Lys	Glu	Asn	Lys	Cys	Leu	Leu	Met	Ser	Phe	Phe	Ile	Leu	Leu	Leu	
				80					85					90	
Ile	Ile	Leu	Leu	Ala	Glu	Val	Thr	Leu	Ala	Ile	Leu	Leu	Phe	Val	
				95					100					105	
Tyr	Glu	Gln	Lys	Leu	Asn	Glu	Tyr	Val	Ala	Lys	Gly	Leu	Thr	Asp	
				110					115					120	
Ser	Ile	His	Arg	Tyr	His	Ser	Asp	Asn	Ser	Thr	Lys	Ala	Ala	Trp	
				125					130					135	
Asp	Ser	Ile	Gln	Ser	Phe	Leu	Gln	Cys	Cys	Gly	Ile	Asn	Gly	Thr	
				140					145					150	
Ser	Asp	Leu	Asp	Ser	Gly	Ser	Pro	Ala	Ser	Cys	Pro	Ser	Asp	Arg	
				155					160					165	
Lys	Val	Glu	Gly	Cys	Tyr	Ala	Lys	Glu	Asp	Phe	Gly	Phe	Ile	Gln	
				170					175					180	
Phe	Pro	Val	Tyr	Arg	Asn	His	His	His	Leu	Cys	Met	Cys	Asp		
				185					190						

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 342 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: BEPINOT01  
(B) CLONE: 546656

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3 :

Met	Ser	Leu	His	Gly	Lys	Arg	Lys	Glu	Ile	Tyr	Lys	Tyr	Glu	Ala	
				5				10						15	
Pro	Trp	Thr	Val	Tyr	Ala	Met	Asn	Trp	Ser	Val	Arg	Pro	Asp	Lys	
				20				25						30	
Arg	Phe	Arg	Leu	Ala	Leu	Gly	Ser	Phe	Val	Glu	Glu	Tyr	Asn	Asn	
				35				40						45	
Lys	Val	Gln	Leu	Val	Gly	Leu	Asp	Glu	Glu	Ser	Ser	Glu	Phe	Ile	
				50				55						60	

PF-0459 US

Cys	Arg	Asn	Thr	Phe	Asp	His	Pro	Tyr	Pro	Thr	Thr	Lys	Leu	Met	
				65					70					75	
Trp	Ile	Pro	Asp	Thr	Lys	Gly	Val	Tyr	Pro	Asp	Leu	Leu	Ala	Thr	
				80					85					90	
Ser	Gly	Asp	Tyr	Leu	Arg	Val	Trp	Arg	Val	Gly	Glu	Thr	Glu	Thr	
				95					100					105	
Arg	Leu	Glu	Cys	Leu	Leu	Asn	Asn	Asn	Lys	Asn	Ser	Asp	Phe	Cys	
				110					115					120	
Ala	Pro	Leu	Thr	Ser	Phe	Asp	Trp	Asn	Glu	Val	Asp	Pro	Tyr	Leu	
				125					130					135	
Leu	Gly	Thr	Ser	Ser	Ile	Asp	Thr	Thr	Cys	Thr	Ile	Trp	Gly	Leu	
				140					145					150	
Glu	Thr	Gly	Gln	Val	Leu	Gly	Arg	Val	Asn	Leu	Val	Ser	Gly	His	
				155					160					165	
Val	Lys	Thr	Gln	Leu	Ile	Ala	His	Asp	Lys	Glu	Val	Tyr	Asp	Ile	
				170					175					180	
Ala	Phe	Ser	Arg	Ala	Gly	Gly	Gly	Arg	Asp	Met	Phe	Ala	Ser	Val	
				185					190					195	
Gly	Ala	Asp	Gly	Ser	Val	Arg	Met	Phe	Asp	Leu	Arg	His	Leu	Glu	
				200					205					210	
His	Ser	Thr	Ile	Ile	Tyr	Glu	Asp	Pro	Gln	His	His	Pro	Leu	Leu	
				215					220					225	
Arg	Leu	Cys	Trp	Asn	Lys	Gln	Asp	Pro	Asn	Tyr	Leu	Ala	Thr	Met	
				230					235					240	
Ala	Met	Asp	Gly	Met	Glu	Val	Val	Ile	Leu	Asp	Val	Arg	Val	Pro	
				245					250					255	
Cys	Thr	Pro	Val	Ala	Arg	Leu	Asn	Asn	His	Arg	Ala	Cys	Val	Asn	
				260					265					270	
Gly	Ile	Ala	Trp	Ala	Pro	His	Ser	Ser	Cys	His	Ile	Cys	Thr	Ala	
				275					280					285	
Ala	Asp	Asp	His	Gln	Ala	Leu	Ile	Trp	Asp	Ile	Gln	Gln	Met	Pro	
				290					295					300	
Arg	Ala	Ile	Glu	Asp	Pro	Ile	Leu	Ala	Tyr	Thr	Ala	Glu	Gly	Glu	
				305					310					315	
Ile	Asn	Asn	Val	Gln	Trp	Ala	Ser	Thr	Gln	Pro	Asp	Trp	Ile	Ala	
				320					325					330	
Ile	Cys	Tyr	Asn	Asn	Cys	Leu	Glu	Ile	Leu	Arg	Val				
				335					340						

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 656 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: SYNORAT03  
(B) CLONE: 693453

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4 :

Met	Glu	Glu	Leu	Asp	Gly	Glu	Pro	Thr	Val	Thr	Leu	Ile	Pro	Gly	
				5					10					15	
Val	Asn	Ser	Lys	Lys	Asn	Gln	Met	Tyr	Phe	Asp	Trp	Gly	Pro	Gly	
				20					25					30	
Glu	Met	Leu	Val	Cys	Glu	Thr	Ser	Phe	Asn	Lys	Lys	Glu	Lys	Ser	

	35	40	45
Glu Met Val Pro Ser Cys Pro Phe Ile Tyr Ile Ile Arg Lys Asp	50	55	60
Val Asp Val Tyr Ser Gln Ile Leu Arg Lys Leu Phe Asn Glu Ser	65	70	75
His Gly Ile Phe Leu Gly Leu Gln Arg Ile Asp Glu Glu Leu Thr	80	85	90
Gly Lys Ser Arg Lys Ser Gln Leu Val Arg Val Ser Lys Asn Tyr	95	100	105
Arg Ser Val Ile Arg Ala Cys Met Glu Glu Met His Gln Val Ala	110	115	120
Ile Ala Ala Lys Asp Pro Ala Asn Gly Arg Gln Phe Ser Ser Gln	125	130	135
Val Ser Ile Leu Ser Ala Met Glu Leu Ile Trp Asn Leu Cys Glu	140	145	150
Ile Leu Phe Ile Glu Val Ala Pro Ala Gly Pro Leu Leu Leu His	155	160	165
Leu Leu Asp Trp Val Arg Leu His Val Cys Glu Val Asp Ser Leu	170	175	180
Ser Ala Asp Val Leu Gly Ser Glu Asn Pro Ser Lys His Asp Ser	185	190	195
Phe Trp Asn Leu Val Thr Ile Leu Val Leu Gln Gly Arg Leu Asp	200	205	210
Glu Ala Arg Gln Met Leu Ser Lys Glu Ala Asp Ala Ser Pro Ala	215	220	225
Ser Ala Gly Ile Cys Arg Ile Met Gly Asp Leu Met Arg Thr Met	230	235	240
Pro Ile Leu Ser Pro Gly Asn Thr Gln Thr Leu Thr Glu Leu Glu	245	250	255
Leu Lys Trp Gln His Trp His Glu Glu Cys Glu Arg Tyr Leu Gln	260	265	270
Asp Ser Thr Phe Ala Thr Ser Pro His Leu Glu Ser Leu Leu Lys	275	280	285
Ile Met Leu Gly Asp Glu Ala Ala Leu Leu Glu Gln Lys Glu Leu	290	295	300
Leu Ser Asn Trp Tyr His Phe Leu Val Thr Arg Leu Leu Tyr Ser	305	310	315
Asn Pro Thr Val Lys Pro Ile Asp Leu His Tyr Tyr Ala Gln Ser	320	325	330
Ser Leu Asp Leu Phe Leu Gly Gly Glu Ser Ser Pro Glu Pro Leu	335	340	345
Asp Asn Ile Leu Leu Ala Ala Phe Glu Phe Asp Ile His Gln Val	350	355	360
Ile Lys Glu Cys Ser Ile Ala Leu Ser Asn Trp Trp Phe Val Ala	365	370	375
His Leu Thr Asp Leu Leu Asp His Cys Lys Leu Leu Gln Ser His	380	385	390
Asn Leu Tyr Phe Gly Ser Asn Met Arg Glu Phe Leu Leu Leu Glu	395	400	405
Tyr Ala Ser Gly Leu Phe Ala His Pro Ser Leu Trp Gln Leu Gly	410	415	420
Val Asp Tyr Phe Asp Tyr Cys Pro Glu Leu Gly Arg Val Ser Leu	425	430	435
Glu Leu His Ile Glu Arg Ile Pro Leu Asn Thr Glu Gln Lys Ala	440	445	450
Leu Lys Val Leu Arg Ile Cys Glu Gln Arg Gln Met Thr Glu Gln	455	460	465
Val Arg Ser Ile Cys Lys Ile Leu Ala Met Lys Ala Val Arg Asn	470	475	480
Asn Arg Leu Gly Ser Ala Leu Ser Trp Ser Ile Arg Ala Lys Asp	485	490	495



Ala	Ala	Phe	Ala	Thr	Leu	Val	Ser	Asp	Arg	Phe	Leu	Arg	Asp	Tyr	
				500					505					510	
Cys	Glu	Arg	Gly	Cys	Phe	Ser	Asp	Leu	Asp	Leu	Ile	Asp	Asn	Leu	
				515					520					525	
Gly	Pro	Ala	Met	Met	Leu	Ser	Asp	Arg	Leu	Thr	Phe	Leu	Gly	Lys	
				530					535					540	
Tyr	Arg	Glu	Phe	His	Arg	Met	Tyr	Gly	Glu	Lys	Arg	Phe	Ala	Asp	
				545					550					555	
Ala	Ala	Ser	Leu	Leu	Leu	Ser	Leu	Met	Thr	Ser	Arg	Ile	Ala	Pro	
				560					565					570	
Arg	Ser	Phe	Trp	Met	Thr	Leu	Leu	Thr	Asp	Ala	Leu	Pro	Leu	Leu	
				575					580					585	
Glu	Gln	Lys	Gln	Val	Ile	Phe	Ser	Ala	Glu	Gln	Thr	Tyr	Glu	Leu	
				590					595					600	
Met	Arg	Cys	Leu	Glu	Asp	Leu	Thr	Ser	Arg	Arg	Pro	Val	His	Gly	
				605					610					615	
Glu	Ser	Asp	Thr	Glu	Gln	Leu	Gln	Asp	Asp	Asp	Ile	Glu	Thr	Thr	
				620					625					630	
Lys	Val	Glu	Met	Leu	Arg	Leu	Ser	Leu	Ala	Arg	Asn	Leu	Ala	Arg	
				635					640					645	
Ala	Ile	Ile	Arg	Glu	Gly	Ser	Leu	Glu	Gly	Ser					
				650					655						

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 236 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAITUT03
- (B) CLONE: 866885

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5 :

Met	Ala	Pro	Asp	Pro	Trp	Phe	Ser	Thr	Tyr	Asp	Ser	Thr	Cys	Gln	
				5					10					15	
Ile	Ala	Gln	Glu	Ile	Ala	Glu	Lys	Ile	Gln	Gln	Arg	Asn	Gln	Tyr	
				20					25					30	
Glu	Arg	Lys	Gly	Glu	Lys	Ala	Pro	Lys	Leu	Thr	Val	Thr	Ile	Arg	
				35					40					45	
Ala	Leu	Leu	Gln	Asn	Leu	Lys	Glu	Lys	Ile	Ala	Leu	Leu	Lys	Asp	
				50					55					60	
Leu	Leu	Leu	Arg	Ala	Val	Ser	Thr	His	Gln	Ile	Thr	Gln	Leu	Glu	
				65					70					75	
Gly	Asp	Arg	Arg	Gln	Asn	Leu	Leu	Asp	Asp	Leu	Val	Thr	Arg	Glu	
				80					85					90	
Arg	Leu	Leu	Leu	Ala	Ser	Phe	Lys	Asn	Glu	Gly	Ala	Glu	Pro	Asp	
				95					100					105	
Leu	Ile	Arg	Ser	Ser	Leu	Met	Ser	Glu	Glu	Ala	Lys	Arg	Gly	Ala	
				110					115					120	
Pro	Asn	Pro	Trp	Leu	Phe	Glu	Glu	Pro	Glu	Glu	Thr	Arg	Gly	Leu	
				125					130					135	
Gly	Phe	Asp	Glu	Ile	Arg	Gln	Gln	Gln	Gln	Lys	Ile	Ile	Gln	Glu	
				140					145					150	
Gln	Asp	Ala	Gly	Leu	Asp	Ala	Leu	Ser	Ser	Ile	Ile	Ser	Arg	Gln	

Lys	Gln	Met	Gly	155	Gln	Glu	Ile	Gly	Asn	160	Glu	Leu	Asp	Glu	Gln	Asn	165
				170						175							180
Glu	Ile	Ile	Asp	185	Asp	Leu	Ala	Asn	Leu	190	Val	Glu	Asn	Thr	Asp	Glu	195
Lys	Leu	Arg	Asn	200	Glu	Thr	Arg	Arg	Val	205	Asn	Met	Val	Asp	Arg	Lys	210
Ser	Ala	Ser	Cys	215	Gly	Met	Ile	Met	Val	220	Ile	Leu	Leu	Leu	Leu	Val	225
Ala	Ile	Val	Val	230	Val	Ala	Val	Trp	Pro	235	Thr	Asn					

## (2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 195 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: LUNGNOT03  
 (B) CLONE: 1242271

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6 :

Met	Leu	Leu	Asp	Thr	Val	Gln	Lys	Val	Phe	Gln	Lys	Met	Leu	Glu			
				5					10					15			
Cys	Ile	Ala	Arg	Ser	Phe	Arg	Lys	Gln	Pro	Glu	Glu	Gly	Leu	Arg			
				20					25					30			
Leu	Leu	Tyr	Ser	Val	Gln	Arg	Pro	Leu	His	Glu	Phe	Ile	Thr	Ala			
				35					40					45			
Val	Gln	Ser	Arg	His	Thr	Asp	Thr	Pro	Val	His	Arg	Gly	Val	Leu			
				50					55					60			
Ser	Thr	Leu	Ile	Ala	Gly	Pro	Val	Val	Glu	Ile	Ser	His	Gln	Leu			
				65					70					75			
Arg	Lys	Val	Ser	Asp	Val	Glu	Glu	Leu	Thr	Pro	Pro	Glu	His	Leu			
				80					85					90			
Ser	Asp	Leu	Pro	Pro	Phe	Ser	Arg	Cys	Leu	Ile	Gly	Ile	Ile	Ile			
				95					100					105			
Lys	Ser	Ser	Asn	Val	Val	Arg	Ser	Phe	Leu	Asp	Glu	Leu	Lys	Ala			
				110					115					120			
Cys	Val	Ala	Ser	Asn	Asp	Ile	Glu	Gly	Ile	Val	Cys	Leu	Thr	Ala			
				125					130					135			
Ala	Val	His	Ile	Ile	Leu	Val	Ile	Asn	Ala	Gly	Lys	His	Lys	Ser			
				140					145					150			
Ser	Lys	Val	Arg	Glu	Val	Ala	Ala	Thr	Val	His	Arg	Lys	Leu	Lys			
				155					160					165			
Thr	Phe	Met	Glu	Ile	Thr	Leu	Glu	Glu	Asp	Ser	Ile	Glu	Arg	Phe			
				170					175					180			
Leu	Tyr	Glu	Ser	Ser	Ser	Arg	Thr	Leu	Gly	Glu	Leu	Leu	Asn	Ser			
				185					190					195			

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 608 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: LUNGFET03  
 (B) CLONE: 1255027

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7 :

Met	Thr	Lys	Thr	Asp	Glu	Thr	Thr	Leu	Val	Ala	Ser	Trp	Glu	Thr	5	10	15
Arg	Glu	Lys	Thr	Ala	Lys	Thr	Thr	Leu	Phe	Leu	Pro	Leu	Glu	Phe	20	25	30
Trp	Ser	Tyr	Lys	Ala	Glu	Val	Pro	His	Leu	Pro	Glu	Leu	Ala	Tyr	35	40	45
Ser	Ala	Arg	Ser	Lys	Met	Ala	Glu	Leu	Asn	Thr	His	Val	Asn	Val	50	55	60
Lys	Glu	Lys	Ile	Tyr	Ala	Val	Arg	Ser	Val	Val	Pro	Asn	Lys	Ser	65	70	75
Asn	Asn	Glu	Ile	Val	Leu	Val	Leu	Gln	Gln	Phe	Asp	Phe	Asn	Val	80	85	90
Asp	Lys	Ala	Val	Gln	Ala	Phe	Val	Asp	Gly	Ser	Ala	Ile	Gln	Val	95	100	105
Leu	Lys	Glu	Trp	Asn	Met	Thr	Gly	Lys	Lys	Lys	Asn	Asn	Lys	Arg	110	115	120
Lys	Arg	Ser	Lys	Ser	Lys	Gln	His	Gln	Gly	Asn	Lys	Asp	Ala	Lys	125	130	135
Asp	Lys	Val	Glu	Arg	Pro	Glu	Ala	Gly	Pro	Leu	Gln	Pro	Gln	Pro	140	145	150
Pro	Gln	Ile	Gln	Asn	Gly	Pro	Met	Asn	Gly	Cys	Glu	Lys	Asp	Ser	155	160	165
Ser	Ser	Thr	Asp	Ser	Ala	Asn	Glu	Lys	Pro	Ala	Leu	Ile	Pro	Arg	170	175	180
Glu	Lys	Lys	Ile	Ser	Ile	Leu	Glu	Glu	Pro	Ser	Lys	Ala	Leu	Arg	185	190	195
Gly	Val	Thr	Glu	Gly	Asn	Arg	Leu	Leu	Gln	Gln	Lys	Leu	Ser	Leu	200	205	210
Asp	Gly	Asn	Pro	Lys	Pro	Ile	His	Gly	Thr	Thr	Glu	Arg	Ser	Asp	215	220	225
Gly	Leu	Gln	Trp	Ser	Ala	Glu	Gln	Pro	Cys	Asn	Pro	Ser	Lys	Pro	230	235	240
Lys	Ala	Lys	Thr	Ser	Pro	Val	Lys	Ser	Asn	Thr	Pro	Ala	Ala	His	245	250	255
Leu	Glu	Ile	Lys	Pro	Asp	Glu	Leu	Ala	Lys	Lys	Arg	Gly	Pro	Asn	260	265	270
Ile	Glu	Lys	Ser	Val	Lys	Asp	Leu	Gln	Arg	Cys	Thr	Val	Ser	Leu	275	280	285
Thr	Arg	Tyr	Arg	Val	Met	Ile	Lys	Glu	Glu	Val	Asp	Ser	Ser	Val	290	295	300
Lys	Lys	Ile	Lys	Ala	Ala	Phe	Ala	Glu	Leu	His	Asn	Cys	Ile	Ile	305	310	315
Asp	Lys	Glu	Val	Ser	Leu	Met	Ala	Glu	Met	Asp	Lys	Val	Lys	Glu	320	325	330
Glu	Ala	Met	Glu	Ile	Leu	Thr	Ala	Arg	Gln	Lys	Lys	Ala	Glu	Glu	335	340	345

PF-0459 US

Leu	Lys	Arg	Leu	Thr	Asp	Leu	Ala	Ser	Gln	Met	Ala	Glu	Met	Gln	350	355	360
Leu	Ala	Glu	Leu	Arg	Ala	Glu	Ile	Lys	His	Phe	Val	Ser	Glu	Arg	365	370	375
Lys	Tyr	Asp	Glu	Glu	Leu	Gly	Lys	Ala	Ala	Arg	Phe	Ser	Cys	Asp	380	385	390
Ile	Glu	Gln	Leu	Lys	Ala	Gln	Ile	Met	Leu	Cys	Gly	Glu	Ile	Thr	395	400	405
His	Pro	Lys	Asn	Asn	Tyr	Ser	Ser	Arg	Thr	Pro	Cys	Ser	Ser	Leu	410	415	420
Leu	Pro	Leu	Leu	Asn	Ala	His	Ala	Ala	Thr	Ser	Gly	Lys	Gln	Ser	425	430	435
Asn	Phe	Ser	Arg	Lys	Ser	Ser	Thr	His	Asn	Lys	Pro	Ser	Glu	Gly	440	445	450
Lys	Ala	Ala	Asn	Pro	Lys	Met	Val	Ser	Ser	Leu	Pro	Ser	Thr	Ala	455	460	465
Asp	Pro	Ser	His	Gln	Thr	Met	Pro	Ala	Asn	Lys	Gln	Asn	Gly	Ser	470	475	480
Ser	Asn	Gln	Arg	Arg	Arg	Phe	Asn	Pro	Gln	Tyr	His	Asn	Asn	Arg	485	490	495
Leu	Asn	Gly	Pro	Ala	Lys	Ser	Gln	Gly	Ser	Gly	Asn	Glu	Ala	Glu	500	505	510
Pro	Leu	Gly	Lys	Gly	Asn	Ser	Arg	His	Glu	His	Arg	Arg	Gln	Pro	515	520	525
His	Asn	Gly	Phe	Arg	Pro	Lys	Asn	Lys	Gly	Gly	Ala	Lys	Asn	Gln	530	535	540
Glu	Ala	Ser	Leu	Gly	Met	Lys	Thr	Pro	Glu	Ala	Pro	Ala	His	Ser	545	550	555
Glu	Lys	Pro	Arg	Arg	Arg	Gln	His	Ala	Ala	Asp	Thr	Ser	Glu	Ala	560	565	570
Arg	Pro	Phe	Arg	Gly	Ser	Val	Gly	Arg	Val	Ser	Gln	Cys	Asn	Leu	575	580	585
Cys	Pro	Thr	Arg	Ile	Glu	Val	Ser	Thr	Asp	Ala	Ala	Val	Leu	Ser	590	595	600
Val	Pro	Ala	Val	Thr	Leu	Val	Ala								605		

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 267 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: TESTTUT02
- (B) CLONE: 1273453

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8 :

Met	Val	Ile	Ser	Trp	His	Leu	Ala	Ser	Asp	Met	Asp	Cys	Val	Val	5	10	15
Thr	Leu	Thr	Thr	Asp	Ala	Ala	Arg	Arg	Ile	Tyr	Asp	Glu	Thr	Gln	20	25	30
Gly	Arg	Gln	Gln	Val	Leu	Pro	Leu	Asp	Ser	Ile	Tyr	Lys	Lys	Thr	35	40	45
Leu	Pro	Asp	Trp	Lys	Arg	Ser	Leu	Pro	His	Phe	Arg	Asn	Gly	Lys			

	50		55		60
Leu Tyr Phe Lys	Pro Ile Gly Asp	Pro Val Phe Ala Arg	Asp Leu		
	65		70		75
Leu Thr Phe Pro	Asp Asn Val Glu His	Cys Glu Thr Val	Phe Gly		
	80		85		90
Met Leu Leu Gly	Asp Thr Ile Ile Leu	Asp Asn Leu Asp	Ala Ala		
	95		100		105
Asn His Tyr Arg	Lys Glu Val Val Lys	Ile Thr His Cys	Pro Thr		
	110		115		120
Leu Leu Thr Arg	Gly Asp Arg Ile	Arg Ser Asn Gly	Lys Phe		
	125		130		135
Gly Gly Leu Gln	Asn Lys Ala Pro	Pro Met Asp Lys	Leu Arg Gly		
	140		145		150
Met Val Phe Gly	Ala Pro Val Pro	Lys Gln Cys Leu	Ile Leu Gly		
	155		160		165
Glu Gln Ile Asp	Leu Leu Gln Gln	Tyr Arg Ser Ala	Val Cys Lys		
	170		175		180
Leu Asp Ser Val	Asn Lys Asp Leu	Asn Ser Gln Leu	Glu Tyr Leu		
	185		190		195
Arg Thr Pro Asp	Met Arg Lys Lys	Lys Gln Glu Leu	Asp Glu His		
	200		205		210
Glu Lys Asn Leu	Lys Leu Ile Glu	Glu Lys Leu Gly	Met Thr Pro		
	215		220		225
Ile Arg Lys Cys	Asn Asp Ser Leu	Arg His Ser Pro	Lys Val Glu		
	230		235		240
Thr Thr Asp Cys	Pro Val Pro Pro	Lys Arg Met Arg	Arg Glu Ala		
	245		250		255
Thr Arg Gln Asn	Arg Ile Ile Thr	Lys Thr Asp Val			
	260		265		

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 285 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: TESTTUT02
- (B) CLONE: 1275261

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9 :

Met Val Met Arg	Pro Leu Trp Ser	Leu Leu Leu	Trp Glu Ala	Leu	
	5		10		15
Leu Pro Ile Thr	Val Thr Gly Ala	Gln Val Leu	Ser Lys Val	Gly	
	20		25		30
Gly Ser Val Leu	Leu Val Ala Ala	Arg Pro Pro	Gly Phe Gln	Val	
	35		40		45
Arg Glu Ala Ile	Trp Arg Ser Leu	Trp Pro Ser	Glu Glu Leu	Leu	
	50		55		60
Ala Thr Phe Phe	Arg Gly Ser Leu	Glu Thr Leu	Tyr His Ser	Arg	
	65		70		75
Phe Leu Gly Arg	Ala Gln Leu His	Ser Asn Leu	Ser Leu Glu	Leu	
	80		85		90
Gly Pro Leu Glu	Ser Gly Asp Ser	Gly Asn Phe	Ser Val Leu	Met	
	95		100		105

PF-0459 US

Val	Asp	Thr	Arg	Gly	Gln	Pro	Trp	Thr	Gln	Thr	Leu	Gln	Leu	Lys
				110					115					120
Val	Tyr	Asp	Ala	Val	Pro	Arg	Pro	Val	Val	Gln	Val	Phe	Ile	Ala
				125					130					135
Val	Glu	Arg	Asp	Ala	Gln	Pro	Ser	Lys	Thr	Cys	Gln	Val	Phe	Leu
				140					145					150
Ser	Cys	Trp	Ala	Pro	Asn	Ile	Ser	Glu	Ile	Thr	Tyr	Ser	Trp	Arg
				155					160					165
Arg	Glu	Thr	Thr	Met	Asp	Phe	Gly	Met	Glu	Pro	His	Ser	Leu	Phe
				170					175					180
Thr	Asp	Gly	Gln	Val	Leu	Ser	Ile	Ser	Leu	Gly	Pro	Gly	Asp	Arg
				185					190					195
Asp	Val	Ala	Tyr	Ser	Cys	Ile	Val	Ser	Asn	Pro	Val	Ser	Trp	Asp
				200					205					210
Leu	Ala	Thr	Val	Thr	Pro	Trp	Asp	Ser	Cys	His	His	Glu	Ala	Ala
				215					220					225
Pro	Gly	Lys	Ala	Ser	Tyr	Lys	Asp	Val	Leu	Leu	Val	Val	Val	Pro
				230					235					240
Val	Ser	Leu	Leu	Leu	Met	Leu	Val	Thr	Leu	Phe	Ser	Ala	Trp	His
				245					250					255
Trp	Cys	Pro	Cys	Ser	Gly	Lys	Lys	Lys	Lys	Asp	Val	His	Ala	Asp
				260					265					270
Arg	Val	Gly	Pro	Glu	Thr	Glu	Asn	Pro	Leu	Val	Gln	Asp	Leu	Pro
				275					280					285

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 76 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: COLNNOT16  
 (B) CLONE: 1281682

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10 :

Met	Pro	Phe	Thr	Arg	Pro	Leu	Lys	His	Phe	Val	Ser	Leu	Leu	His
				5					10					15
Pro	Ser	Ala	Ser	Gln	Val	His	Asn	Ala	Gly	Gln	His	Gln	Lys	Leu
				20					25					30
Lys	Thr	Leu	Glu	Lys	Ala	Cys	Gly	Leu	Ala	Leu	Gly	Glu	Gly	Arg
				35					40					45
Glu	Gln	Asn	Leu	Cys	Thr	Ser	Leu	Phe	Asn	Leu	Glu	Ile	Arg	His
				50					55					60
Pro	Arg	Asp	Ala	Ile	Ile	Phe	Cys	Val	Ser	Ile	Val	Val	Pro	Leu
				65					70					75
Ser														

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 147 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: BRSTNOT07  
 (B) CLONE: 1298305

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11 :

Met	Thr	Ala	Ser	Thr	Gly	His	Leu	Gly	Leu	Gly	Trp	Ser	Ala	Arg	
				5					10					15	
Pro	Cys	Pro	Cys	Gly	Thr	Leu	Gly	Ser	Cys	Phe	Leu	Ser	Leu	Phe	
				20					25					30	
Ala	Ala	Leu	Leu	Trp	Leu	Ala	Ala	Ala	Val	Leu	Gln	Ala	Cys	Val	
				35					40					45	
Gly	His	Ser	Asp	Glu	Gly	Cys	Gly	Ala	Ser	Gln	Cys	Arg	Arg	Ala	
				50					55					60	
Ala	Leu	Gly	Ile	Val	Pro	Ser	Pro	Val	Ser	Val	Leu	Arg	Thr	Tyr	
				65					70					75	
Pro	Gly	Leu	His	His	Gln	Asp	Pro	Val	Phe	Gly	Phe	Arg	Arg	Pro	
				80					85					90	
Ser	Met	Gly	Lys	Thr	Arg	His	Gln	Pro	Leu	Gln	Gln	Trp	Val	Pro	
				95					100					105	
Leu	Ala	Cys	Gly	His	Gln	Leu	Gly	Asp	Pro	Gly	Ser	Gly	Pro	Leu	
				110					115					120	
Leu	Ser	Pro	Val	Ser	Leu	Cys	Cys	Gly	Phe	Trp	Ala	Val	Met	Ser	
				125					130					135	
Pro	Pro	Leu	Lys	Asp	Val	Phe	Thr	Leu	Thr	Ser	Gly				
				140					145						

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 261 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: LUNGNOT12  
 (B) CLONE: 1360501

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12 :

Met	Glu	Leu	Leu	Gln	Val	Thr	Ile	Leu	Phe	Leu	Leu	Pro	Ser	Ile	
				5					10					15	
Cys	Ser	Ser	Asn	Ser	Thr	Gly	Val	Leu	Glu	Ala	Ala	Asn	Asn	Ser	
				20					25					30	
Leu	Val	Val	Thr	Thr	Thr	Lys	Pro	Ser	Ile	Thr	Thr	Pro	Asn	Thr	
				35					40					45	
Glu	Ser	Leu	Gln	Lys	Asn	Val	Val	Thr	Pro	Thr	Thr	Gly	Thr	Thr	
				50					55					60	
Pro	Lys	Gly	Thr	Ile	Thr	Asn	Glu	Leu	Leu	Lys	Met	Ser	Leu	Met	
				65					70					75	
Ser	Thr	Ala	Thr	Phe	Leu	Thr	Ser	Lys	Asp	Glu	Gly	Leu	Lys	Ala	
				80					85					90	

PF-0459 US

Thr	Thr	Thr	Asp	Val	Arg	Lys	Asn	Asp	Ser	Ile	Ile	Ser	Asn	Val	
				95					100					105	
Thr	Val	Thr	Ser	Val	Thr	Leu	Pro	Asn	Ala	Val	Ser	Thr	Leu	Gln	
				110					115					120	
Ser	Ser	Lys	Pro	Lys	Thr	Glu	Thr	Gln	Ser	Ser	Ile	Lys	Thr	Thr	
				125					130					135	
Glu	Ile	Pro	Gly	Ser	Val	Leu	Gln	Pro	Asp	Ala	Ser	Pro	Ser	Lys	
				140					145					150	
Thr	Gly	Thr	Leu	Thr	Ser	Ile	Pro	Val	Thr	Ile	Pro	Glu	Asn	Thr	
				155					160					165	
Ser	Gln	Ser	Gln	Val	Ile	Gly	Thr	Glu	Gly	Gly	Lys	Asn	Ala	Ser	
				170					175					180	
Thr	Ser	Ala	Thr	Ser	Arg	Ser	Tyr	Ser	Ser	Ile	Ile	Leu	Pro	Val	
				185					190					195	
Val	Ile	Ala	Leu	Ile	Val	Ile	Thr	Leu	Ser	Val	Phe	Val	Leu	Val	
				200					205					210	
Gly	Leu	Tyr	Arg	Met	Cys	Trp	Lys	Ala	Asp	Pro	Gly	Thr	Pro	Glu	
				215					220					225	
Asn	Gly	Asn	Asp	Gln	Pro	Gln	Ser	Asp	Lys	Glu	Ser	Val	Lys	Leu	
				230					235					240	
Leu	Thr	Val	Lys	Thr	Ile	Ser	His	Glu	Ser	Gly	Glu	His	Ser	Ala	
				245					250					255	
Gln	Gly	Lys	Thr	Lys	Asn										
				260											

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 213 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGNOT12
- (B) CLONE: 1362406

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13 :

Met	Ala	Gly	Cys	Pro	Ala	Asp	Arg	Ser	Ile	Leu	Ala	Pro	Leu	Ala	
				5					10					15	
Trp	Asp	Leu	Gly	Leu	Leu	Leu	Leu	Phe	Val	Gly	Gln	His	Ser	Leu	
				20					25					30	
Met	Ala	Ala	Glu	Arg	Val	Lys	Ala	Trp	Thr	Ser	Arg	Tyr	Phe	Gly	
				35					40					45	
Val	Leu	Gln	Arg	Ser	Leu	Tyr	Val	Ala	Cys	Thr	Ala	Leu	Ala	Leu	
				50					55					60	
Gln	Leu	Val	Met	Arg	Tyr	Trp	Glu	Pro	Ile	Pro	Lys	Gly	Pro	Val	
				65					70					75	
Leu	Trp	Glu	Ala	Arg	Ala	Glu	Pro	Trp	Ala	Thr	Trp	Val	Pro	Leu	
				80					85					90	
Leu	Cys	Phe	Val	Leu	His	Val	Ile	Ser	Trp	Leu	Leu	Ile	Phe	Ser	
				95					100					105	
Ile	Leu	Leu	Val	Phe	Asp	Tyr	Ala	Glu	Leu	Met	Gly	Leu	Lys	Gln	
				110					115					120	
Val	Tyr	Tyr	His	Val	Leu	Gly	Leu	Gly	Glu	Pro	Leu	Ala	Leu	Lys	
				125					130					135	
Ser	Pro	Arg	Ala	Leu	Arg	Leu	Phe	Ser	His	Leu	Arg	His	Pro	Val	



PF-0459 US

	140		145		150
Cys Val Glu Leu	Leu Thr Val Leu Trp	Val Val Pro Thr Leu	Gly		
	155		160		165
Thr Asp Arg Leu	Leu Leu Ala Phe Leu	Leu Thr Leu Tyr Leu	Gly		
	170		175		180
Leu Ala His Gly	Leu Asp Gln Gln Asp	Leu Arg Tyr Leu Arg	Ala		
	185		190		195
Gln Leu Gln Arg	Lys Leu His Leu Leu	Ser Arg Pro Gln Asp	Gly		
	200		205		210
Glu Ala Glu					

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 67 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: LATRTUT02  
(B) CLONE: 1405329

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14 :

Met Gln Pro Arg	Pro Arg Gly Arg	Pro Pro Arg Thr Arg	Gly Asp
	5	10	15
Glu Ala Pro Gln	Trp His Leu Pro Asp	Ala Ala Ala Leu Leu	Pro
	20	25	30
Val Arg Leu Pro	Leu Ala Val Leu Val	Arg Gly Thr Gln Arg	Pro
	35	40	45
Glu Arg Arg Arg	Cys Gly Arg Leu Pro	Ala Gly Val Pro Gly	Ala
	50	55	60
Ala Arg Ser Val	Ala Arg Ser		
	65		

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 161 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: BRAINOT12  
(B) CLONE: 1415223

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15 :

Met Leu Ala Pro	Gln Arg Thr Arg	Ala Pro Ser Pro	Arg Ala Ala
	5	10	15
Pro Arg Pro Thr	Arg Ser Met Leu Pro	Ala Ala Met Lys	Gly Leu
	20	25	30

PF-0459 US

Gly	Leu	Ala	Leu	Leu	Ala	Val	Leu	Leu	Cys	Ser	Ala	Pro	Ala	His	
			35						40					45	
Gly	Leu	Trp	Cys	Gln	Asp	Cys	Thr	Leu	Thr	Thr	Asn	Ser	Ser	His	
			50						55					60	
Cys	Thr	Pro	Lys	Gln	Cys	Gln	Pro	Ser	Asp	Thr	Val	Cys	Ala	Ser	
			65						70					75	
Val	Arg	Ile	Thr	Asp	Pro	Ser	Ser	Ser	Arg	Lys	Asp	His	Ser	Val	
			80						85					90	
Asn	Lys	Met	Cys	Ala	Ser	Ser	Cys	Asp	Phe	Val	Lys	Arg	His	Phe	
			95						100					105	
Phe	Ser	Asp	Tyr	Leu	Met	Gly	Phe	Ile	Asn	Ser	Gly	Ile	Leu	Lys	
			110						115					120	
Val	Asp	Val	Asp	Cys	Cys	Glu	Lys	Asp	Leu	Cys	Asn	Gly	Ala	Ala	
			125						130					135	
Gly	Ala	Gly	His	Ser	Pro	Trp	Ala	Leu	Ala	Gly	Gly	Leu	Leu	Leu	
			140						145					150	
Ser	Leu	Gly	Pro	Ala	Leu	Leu	Trp	Ala	Gly	Pro					
			155						160						

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 141 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAINOT12
- (B) CLONE: 1416553

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16 :

Met	Trp	Ala	Gln	Arg	Val	Leu	Thr	Leu	Trp	Gln	Gly	Leu	Ser	Trp	
				5					10					15	
Gly	Arg	Pro	Pro	Ser	Gly	Pro	Gly	Ala	Met	Ala	Pro	Arg	Gly	Gln	
			20						25					30	
Ala	Asp	Leu	Leu	Pro	Ala	Val	Ser	Thr	Pro	Phe	Leu	Ile	Thr	Val	
			35						40					45	
Trp	Ser	Pro	Ser	Phe	Gly	Cys	Ser	Leu	Arg	Cys	Val	Leu	Gly	Ser	
			50						55					60	
Ser	Glu	Pro	Glu	Ala	Ser	Phe	Trp	Lys	Pro	Ala	Val	Leu	Pro	Ala	
			65						70					75	
Pro	Val	Gln	Lys	Pro	Leu	Ser	Pro	Ala	Phe	Pro	Gln	Ala	Gly	Val	
			80						85					90	
Gly	Val	Gly	Gly	Leu	Cys	Pro	Ser	Ser	Leu	Thr	Leu	Glu	Arg	Trp	
			95						100					105	
Glu	Ala	Gly	Asn	Leu	His	Leu	Gly	Ala	Trp	Ala	Pro	Pro	Leu	Cys	
			110						115					120	
Ala	Ser	Gly	Phe	Pro	Ala	Pro	Gly	Arg	Gly	Cys	Ser	Pro	Ser	Trp	
			125						130					135	
Thr	Pro	Ala	Cys	Pro	Ser										
			140												

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 152 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: KIDNNOT09
- (B) CLONE: 1418517

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17 :

Met	Glu	Asp	Glu	Glu	Val	Ala	Glu	Ser	Trp	Glu	Glu	Ala	Ala	Asp	5	10	15
Ser	Gly	Glu	Ile	Asp	Arg	Arg	Leu	Glu	Lys	Lys	Leu	Lys	Ile	Thr	20	25	30
Gln	Lys	Glu	Ser	Arg	Lys	Ser	Lys	Ser	Pro	Pro	Lys	Val	Pro	Ile	35	40	45
Val	Ile	Gln	Asp	Asp	Ser	Leu	Pro	Ala	Gly	Pro	Pro	Pro	Gln	Ile	50	55	60
Arg	Ile	Leu	Lys	Arg	Pro	Thr	Ser	Asn	Gly	Val	Val	Ser	Ser	Pro	65	70	75
Asn	Ser	Thr	Ser	Arg	Pro	Thr	Leu	Pro	Val	Lys	Ser	Leu	Ala	Gln	80	85	90
Arg	Glu	Ala	Glu	Tyr	Ala	Glu	Ala	Arg	Lys	Arg	Ile	Leu	Gly	Ser	95	100	105
Ala	Ser	Pro	Glu	Glu	Glu	Gln	Glu	Lys	Pro	Ile	Leu	Asp	Arg	Pro	110	115	120
Thr	Arg	Ile	Ser	Gln	Pro	Glu	Asp	Ser	Arg	Gln	Pro	Asn	Asn	Val	125	130	135
Ile	Arg	Gln	Pro	Leu	Gly	Pro	Asp	Gly	Ser	Gln	Gly	Phe	Lys	Gln	140	145	150
Arg	Arg																

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 742 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PANCNOT08
- (B) CLONE: 1438165

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18 :

Met	Ala	Ser	Val	His	Glu	Ser	Leu	Tyr	Phe	Asn	Pro	Met	Met	Thr	5	10	15
Asn	Gly	Val	Val	His	Ala	Asn	Val	Phe	Gly	Ile	Lys	Asp	Trp	Val	20	25	30
Thr	Pro	Tyr	Lys	Ile	Ala	Val	Leu	Val	Leu	Leu	Asn	Glu	Met	Ser	35	40	45
Arg	Thr	Gly	Glu	Gly	Ala	Val	Ser	Leu	Met	Glu	Arg	Arg	Arg	Leu	50	55	60
Asn	Gln	Leu	Leu	Leu	Pro	Leu	Leu	Gln	Gly	Pro	Asp	Ile	Thr	Leu			

	65		70		75
Ser Lys Leu Tyr Lys	Leu Ile Glu Glu Ser	Cys Pro Gln Leu Ala			
	80		85		90
Asn Ser Val Gln Ile	Arg Ile Lys Leu Met	Ala Glu Gly Glu Leu			
	95		100		105
Lys Asp Met Glu Gln	Phe Phe Asp Asp Leu	Ser Asp Ser Phe Ser			
	110		115		120
Gly Thr Glu Pro Glu	Val His Lys Thr Ser	Val Val Gly Leu Phe			
	125		130		135
Leu Arg His Met Ile	Leu Ala Tyr Ser Lys	Leu Ser Phe Ser Gln			
	140		145		150
Val Phe Lys Leu Tyr	Thr Ala Leu Gln Gln	Tyr Phe Gln Asn Gly			
	155		160		165
Glu Lys Lys Thr Val	Glu Asp Ala Asp Met	Glu Leu Thr Ser Arg			
	170		175		180
Asp Glu Gly Glu Arg	Lys Met Glu Lys Glu	Glu Leu Asp Val Ser			
	185		190		195
Val Arg Glu Glu Glu	Val Ser Cys Ser Gly	Pro Leu Ser Gln Lys			
	200		205		210
Gln Ala Glu Phe Phe	Leu Ser Gln Gln Ala	Ser Leu Leu Lys Asn			
	215		220		225
Asp Glu Thr Lys Ala	Leu Thr Pro Ala Ser	Leu Gln Lys Glu Leu			
	230		235		240
Asn Asn Leu Leu Lys	Phe Asn Pro Asp Phe	Ala Glu Ala His Tyr			
	245		250		255
Leu Ser Tyr Leu Asn	Asn Leu Arg Val Gln	Asp Val Phe Ser Ser			
	260		265		270
Thr His Ser Leu Leu	His Tyr Phe Asp Arg	Leu Ile Leu Thr Gly			
	275		280		285
Ala Glu Ser Lys Ser	Asn Gly Glu Glu Gly	Tyr Gly Arg Ser Leu			
	290		295		300
Arg Tyr Ala Ala Leu	Asn Leu Ala Ala Leu	His Cys Arg Phe Gly			
	305		310		315
His Tyr Gln Gln Ala	Glu Leu Ala Leu Gln	Glu Ala Ile Arg Ile			
	320		325		330
Ala Gln Glu Ser Asn	Asp His Val Cys Leu	Gln His Cys Leu Ser			
	335		340		345
Trp Leu Tyr Val Leu	Gly Gln Lys Arg Ser	Asp Ser Tyr Val Leu			
	350		355		360
Leu Glu His Ser Val	Lys Lys Ala Val His	Phe Gly Leu Pro Arg			
	365		370		375
Ala Phe Ala Gly Lys	Thr Ala Asn Lys Leu	Met Asp Ala Leu Lys			
	380		385		390
Asp Ser Asp Leu Leu	His Trp Lys His Ser	Leu Ser Glu Leu Ile			
	395		400		405
Asp Ile Ser Ile Ala	Gln Lys Thr Ala Ile	Trp Arg Leu Tyr Gly			
	410		415		420
Arg Ser Thr Met Ala	Leu Gln Gln Ala Gln	Met Leu Leu Ser Met			
	425		430		435
Asn Ser Leu Glu Ala	Val Asn Ala Gly Val	Gln Gln Asn Asn Thr			
	440		445		450
Glu Ser Phe Ala Val	Ala Leu Cys His Leu	Ala Glu Leu His Ala			
	455		460		465
Glu Gln Gly Cys Phe	Ala Ala Ala Ser Glu	Val Leu Lys His Leu			
	470		475		480
Lys Glu Arg Phe Pro	Pro Asn Ser Gln His	Ala Gln Leu Trp Met			
	485		490		495
Leu Cys Asp Gln Lys	Ile Gln Phe Asp Arg	Ala Met Asn Asp Gly			
	500		505		510
Lys Tyr His Leu Ala	Asp Ser Leu Val Thr	Gly Ile Thr Ala Leu			
	515		520		525

PF-0459 US

Asn Ser Ile Glu Gly Val Tyr Arg Lys Ala Val Val Leu Gln Ala	530	535	540
Gln Asn Gln Met Ser Glu Ala His Lys Leu Leu Gln Lys Leu Leu	545	550	555
Val His Cys Gln Lys Leu Lys Asn Thr Glu Met Val Ile Ser Val	560	565	570
Leu Leu Ser Val Ala Glu Leu Tyr Trp Arg Ser Ser Ser Pro Thr	575	580	585
Ile Ala Leu Pro Met Leu Leu Gln Ala Leu Ala Leu Ser Lys Glu	590	595	600
Tyr Arg Leu Gln Tyr Leu Ala Ser Glu Thr Val Leu Asn Leu Ala	605	610	615
Phe Ala Gln Leu Ile Leu Gly Ile Pro Glu Gln Ala Leu Ser Leu	620	625	630
Leu His Met Ala Ile Glu Pro Ile Leu Ala Asp Gly Ala Ile Leu	635	640	645
Asp Lys Gly Arg Ala Met Phe Leu Val Ala Lys Cys Gln Val Ala	650	655	660
Ser Ala Ala Ser Tyr Asp Gln Pro Lys Lys Ala Glu Ala Leu Glu	665	670	675
Ala Ala Ile Glu Asn Leu Asn Glu Ala Lys Asn Tyr Phe Ala Lys	680	685	690
Val Asp Cys Lys Glu Arg Ile Arg Asp Val Val Tyr Phe Gln Ala	695	700	705
Arg Leu Tyr His Thr Leu Gly Lys Thr Gln Glu Arg Asn Arg Cys	710	715	720
Ala Met Leu Phe Arg Gln Leu His Gln Glu Leu Pro Ser His Gly	725	730	735
Val Pro Leu Ile Asn His Leu	740		

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 805 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: THYRNOT03
- (B) CLONE: 1440381

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19 :

Met Asp Gly Ile Leu Asp Glu Ser Leu Leu Glu Thr Cys Pro Ile	5	10	15
Gln Ser Pro Leu Gln Val Phe Ala Gly Met Gly Gly Leu Ala Leu	20	25	30
Ile Ala Glu Arg Leu Pro Met Leu Tyr Pro Glu Val Ile Gln Gln	35	40	45
Val Ser Ala Pro Val Val Thr Ser Thr Thr Gln Glu Lys Pro Tyr	50	55	60
Asp Ser Asp Gln Phe Glu Trp Val Thr Ile Glu Gln Ser Gly Glu	65	70	75
Leu Val Tyr Glu Ala Pro Glu Thr Val Ala Ala Glu Pro Pro Pro	80	85	90
Ile Lys Ser Ala Val Gln Thr Met Ser Pro Ile Pro Ala His Ser			

	95		100		105									
Leu	Ala	Ala	Phe	Gly	Leu	Phe	Leu	Arg	Leu	Pro	Gly	Tyr	Ala	Glu
	110				115				120					
Val	Leu	Leu	Lys	Glu	Arg	Lys	His	Ala	Gln	Cys	Leu	Leu	Arg	Leu
	125				130				135					
Val	Leu	Gly	Val	Thr	Asp	Asp	Gly	Glu	Gly	Ser	His	Ile	Leu	Gln
	140				145				150					
Ser	Pro	Ser	Ala	Asn	Val	Leu	Pro	Thr	Leu	Pro	Phe	His	Val	Leu
	155				160				165					
Arg	Ser	Leu	Phe	Ser	Thr	Thr	Pro	Leu	Thr	Thr	Asp	Asp	Gly	Val
	170				175				180					
Leu	Leu	Arg	Arg	Met	Ala	Leu	Glu	Ile	Gly	Ala	Leu	His	Leu	Ile
	185				190				195					
Leu	Val	Cys	Leu	Ser	Ala	Leu	Ser	His	His	Ser	Pro	Arg	Val	Pro
	200				205				210					
Asn	Ser	Ser	Val	Asn	Gln	Thr	Glu	Pro	Gln	Val	Ser	Ser	Ser	His
	215				220				225					
Asn	Pro	Thr	Ser	Thr	Glu	Glu	Gln	Gln	Leu	Tyr	Trp	Ala	Lys	Gly
	230				235				240					
Thr	Gly	Phe	Gly	Thr	Gly	Ser	Thr	Ala	Ser	Gly	Trp	Asp	Val	Glu
	245				250				255					
Gln	Ala	Leu	Thr	Lys	Gln	Arg	Leu	Glu	Glu	Glu	His	Val	Thr	Cys
	260				265				270					
Leu	Leu	Gln	Val	Leu	Ala	Ser	Tyr	Ile	Asn	Pro	Val	Ser	Ser	Ala
	275				280				285					
Val	Asn	Gly	Glu	Ala	Gln	Ser	Ser	His	Glu	Thr	Arg	Gly	Gln	Asn
	290				295				300					
Ser	Asn	Ala	Leu	Pro	Ser	Val	Leu	Leu	Glu	Leu	Leu	Ser	Gln	Ser
	305				310				315					
Cys	Leu	Ile	Pro	Ala	Met	Ser	Ser	Tyr	Leu	Arg	Asn	Asp	Ser	Val
	320				325				330					
Leu	Asp	Met	Ala	Arg	His	Val	Pro	Leu	Tyr	Arg	Ala	Leu	Leu	Glu
	335				340				345					
Leu	Leu	Arg	Ala	Ile	Ala	Ser	Cys	Ala	Ala	Met	Val	Pro	Leu	Leu
	350				355				360					
Leu	Pro	Leu	Ser	Thr	Glu	Asn	Gly	Glu	Glu	Glu	Glu	Gln	Ser	
	365				370				375					
Glu	Cys	Gln	Thr	Ser	Val	Gly	Thr	Leu	Leu	Ala	Lys	Met	Lys	Thr
	380				385				390					
Cys	Val	Asp	Thr	Tyr	Thr	Asn	Arg	Leu	Arg	Ser	Lys	Arg	Glu	Asn
	395				400				405					
Val	Lys	Thr	Gly	Val	Lys	Pro	Asp	Ala	Ser	Asp	Gln	Glu	Pro	Glu
	410				415				420					
Gly	Leu	Thr	Leu	Leu	Val	Pro	Asp	Ile	Gln	Lys	Thr	Ala	Glu	Ile
	425				430				435					
Val	Tyr	Ala	Ala	Thr	Thr	Ser	Leu	Arg	Gln	Ala	Asn	Gln	Glu	Lys
	440				445				450					
Asn	Trp	Val	Asn	Thr	Pro	Arg	Arg	Arg	Leu	Met	Asn	Pro	Lys	Pro
	455				460				465					
Leu	Ser	Val	Leu	Lys	Ser	Leu	Glu	Glu	Lys	Tyr	Val	Ala	Val	Met
	470				475				480					
Lys	Lys	Leu	Gln	Phe	Asp	Thr	Phe	Glu	Met	Val	Ser	Glu	Asp	Glu
	485				490				495					
Asp	Gly	Lys	Leu	Gly	Phe	Lys	Val	Asn	Tyr	His	Tyr	Met	Ser	Gln
	500				505				510					
Val	Lys	Asn	Ala	Asn	Asp	Ala	Asn	Ser	Ala	Ala	Arg	Ala	Arg	Arg
	515				520				525					
Leu	Ala	Gln	Glu	Ala	Val	Thr	Leu	Ser	Thr	Ser	Leu	Pro	Leu	Ser
	530				535				540					
Ser	Ser	Ser	Ser	Val	Phe	Val	Arg	Cys	Asp	Glu	Glu	Arg	Leu	Asp
	545				550				555					

PF-0459 US

Ile Met Lys Val	Leu Ile Thr Gly Pro	Ala Asp Thr Pro Tyr	Ala
	560	565	570
Asn Gly Cys Phe	Glu Phe Asp Val Tyr	Phe Pro Gln Asp Tyr	Pro
	575	580	585
Ser Ser Pro Pro	Leu Val Asn Leu Glu	Thr Thr Gly Gly His	Ser
	590	595	600
Val Arg Phe Asn	Pro Asn Leu Tyr Asn	Asp Gly Lys Val Cys	Leu
	605	610	615
Ser Ile Leu Asn	Thr Trp His Gly Arg	Pro Glu Glu Lys Trp	Asn
	620	625	630
Pro Gln Thr Ser	Ser Phe Leu Gln Val	Leu Val Ser Val Gln	Ser
	635	640	645
Leu Ile Leu Val	Ala Glu Pro Tyr Phe	Asn Glu Pro Gly Tyr	Glu
	650	655	660
Arg Ser Arg Gly	Thr Pro Ser Gly Thr	Gln Ser Ser Arg Glu	Tyr
	665	670	675
Asp Gly Asn Ile	Arg Gln Ala Thr Val	Lys Trp Ala Met Leu	Glu
	680	685	690
Gln Ile Arg Asn	Pro Ser Pro Cys Phe	Lys Glu Val Ile His	Lys
	695	700	705
His Phe Tyr Leu	Lys Arg Val Glu Ile	Met Ala Gln Cys Glu	Glu
	710	715	720
Trp Ile Ala Asp	Ile Gln Gln Tyr Ser	Ser Asp Lys Arg Val	Gly
	725	730	735
Arg Thr Met Ser	His His Ala Ala Ala	Leu Lys Arg His Thr	Ala
	740	745	750
Gln Leu Arg Glu	Glu Leu Leu Lys Leu	Pro Cys Pro Glu Gly	Leu
	755	760	765
Asp Pro Asp Thr	Asp Asp Ala Pro Glu	Val Cys Arg Ala Thr	Thr
	770	775	780
Gly Ala Glu Glu	Thr Leu Met His Asp	Gln Val Lys Pro Ser	Ser
	785	790	795
Ser Lys Glu Leu	Pro Ser Asp Phe Gln	Leu	
	800	805	

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 195 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGNOT14
- (B) CLONE: 1510839

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20 :

Met Lys Ala Ser	Gln Cys Cys Cys Cys	Leu Ser His Leu Leu	Ala
	5	10	15
Ser Val Leu Leu	Leu Leu Leu Pro	Glu Leu Ser Gly Pro	Leu
	20	25	30
Ala Val Leu Leu	Gln Ala Ala Glu Ala	Ala Pro Gly Leu Gly	Pro
	35	40	45
Pro Asp Pro Arg	Pro Arg Thr Leu Pro	Pro Leu Pro Pro Gly	Pro
	50	55	60
Thr Pro Ala Gln	Gln Pro Gly Arg Gly	Leu Ala Glu Ala Ala	Gly

Pro Arg Gly Ser	65	Gly Gly Asn Gly	70	Ser Asn Pro Val Ala Gly	75
Leu Glu Thr Asp	80	Asp His Gly Gly Lys	85	Ala Gly Glu Gly Ser Val	90
Gly Gly Gly Leu	95	Ala Val Ser Pro Asn	100	Pro Gly Asp Lys Pro Met	105
Thr Gln Arg Ala	110	Leu Thr Val Leu Met	115	Val Val Ser Gly Ala Val	120
Leu Val Tyr Phe	125	Val Val Arg Thr Val	130	Arg Met Arg Arg Arg Asn	135
Arg Lys Thr Arg	140	Arg Tyr Gly Val Leu	145	Asp Thr Asn Ile Glu Asn	150
Met Glu Leu Thr	155	Pro Leu Glu Gln Asp	160	Asp Glu Asp Asp Asp Asn	165
Thr Leu Phe Asp	170	Ala Asn His Pro Arg	175	Arg Arg Glu Cys Ala Phe	180
	185		190		195

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 161 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: SPLNNOT04  
 (B) CLONE: 1534876

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21 :

Met Trp Phe Leu Gly Cys Thr Gly Pro Gly Cys Gly Cys Ala Gly	5	10	15
Val Cys Lys Val Val Pro Cys Ile Ser Thr Gly Phe Glu Thr Ser	20	25	30
Gly Pro Cys Pro Ser Ser Arg Glu Gly Phe Leu Phe Phe Leu Thr	35	40	45
Gln Val Thr Phe Gln Pro Phe Gln Phe Pro Ser Phe Ser Ala Leu	50	55	60
Pro Ser Asn Ser Ala Asn Pro Gly Val Gly Ser Gln Gly Gly Arg	65	70	75
Glu Cys Pro Thr Thr Phe Ser Gly Gln Pro Leu Thr Pro Lys Pro	80	85	90
Leu Pro Pro Ser Ile Leu His Pro Leu Pro Ile Gln Pro Lys Cys	95	100	105
Pro Gln Leu Gly Leu Ser Cys Ile Pro Val Glu Gly Pro Leu Pro	110	115	120
Cys Leu Ser Glu Val Arg Leu Cys Cys Val Met Gly Arg Leu Cys	125	130	135
Pro Ser Pro Pro Leu Ala Arg Cys Thr Cys Phe Leu Val Cys Thr	140	145	150
Arg Cys Pro Gly Gly Pro Ser Leu Pro Cys Gln	155	160	



## (2) INFORMATION FOR SEQ ID NO: 22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 160 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: SPLNNOT04
- (B) CLONE: 1559131

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22 :

Met	Asp	Lys	Leu	Lys	Lys	Val	Leu	Ser	Gly	Gln	Asp	Thr	Glu	Asp		
				5					10					15		
Arg	Ser	Gly	Leu	Ser	Glu	Val	Val	Glu	Ala	Ser	Ser	Leu	Ser	Trp		
				20					25					30		
Ser	Thr	Arg	Ile	Lys	Gly	Phe	Ile	Ala	Cys	Phe	Ala	Ile	Gly	Ile		
				35					40					45		
Leu	Cys	Ser	Leu	Leu	Gly	Thr	Val	Leu	Leu	Trp	Val	Pro	Arg	Lys		
				50					55					60		
Gly	Leu	His	Leu	Phe	Ala	Val	Phe	Tyr	Thr	Phe	Gly	Asn	Ile	Ala		
				65					70					75		
Ser	Ile	Gly	Ser	Thr	Ile	Phe	Leu	Met	Gly	Pro	Val	Lys	Gln	Leu		
				80					85					90		
Lys	Arg	Met	Phe	Glu	Pro	Thr	Arg	Leu	Ile	Ala	Thr	Ile	Met	Val		
				95					100					105		
Leu	Leu	Cys	Phe	Ala	Leu	Thr	Leu	Cys	Ser	Ala	Phe	Trp	Trp	His		
				110					115					120		
Asn	Lys	Gly	Leu	Ala	Leu	Ile	Phe	Cys	Ile	Leu	Gln	Ser	Leu	Ala		
				125					130					135		
Leu	Thr	Trp	Tyr	Ser	Leu	Ser	Phe	Ile	Pro	Phe	Ala	Arg	Asp	Ala		
				140					145					150		
Val	Lys	Lys	Cys	Phe	Ala	Val	Cys	Leu	Ala							
				155					160							

## (2) INFORMATION FOR SEQ ID NO: 23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BLADNOT03
- (B) CLONE: 1601473

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23 :

Met	Gln	Ala	Lys	Tyr	Ser	Ser	Thr	Arg	Asp	Met	Leu	Asp	Asp	Asp		
				5					10					15		
Gly	Asp	Thr	Thr	Met	Ser	Leu	His	Ser	Gln	Ala	Ser	Ala	Thr	Thr		
				20					25					30		
Arg	His	Pro	Glu	Pro	Arg	Arg	Thr	Glu	His	Arg	Ala	Pro	Ser	Ser		
				35					40					45		
Thr	Trp	Arg	Pro	Val	Ala	Leu	Thr	Leu	Leu	Thr	Leu	Cys	Leu	Val		
				50					55					60		

PF-0459 US

Leu Leu Ile Gly Leu Ala Ala Leu Gly Leu Leu Cys Lys Ser Ala  
65 75  
Leu

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 336 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAITUT12
- (B) CLONE: 1615809

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24 :

Met	Ile	Ser	Tyr	Ile	Val	Leu	Leu	Ser	Ile	Leu	Leu	Trp	Pro	Leu	
				5					10						15
Val	Val	Tyr	His	Glu	Leu	Ile	Gln	Arg	Met	Tyr	Thr	Arg	Leu	Glu	
				20					25						30
Pro	Leu	Leu	Met	Gln	Leu	Asp	Tyr	Ser	Met	Lys	Ala	Glu	Ala	Asn	
				35					40						45
Ala	Leu	His	His	Lys	His	Asp	Lys	Arg	Lys	Arg	Gln	Gly	Lys	Asn	
				50					55						60
Ala	Pro	Pro	Gly	Gly	Asp	Glu	Pro	Leu	Ala	Glu	Thr	Glu	Ser	Glu	
				65					70						75
Ser	Glu	Ala	Glu	Leu	Ala	Gly	Phe	Ser	Pro	Val	Val	Asp	Val	Lys	
				80					85						90
Lys	Thr	Ala	Leu	Ala	Leu	Ala	Ile	Thr	Asp	Ser	Glu	Leu	Ser	Asp	
				95					100						105
Glu	Glu	Ala	Ser	Ile	Leu	Glu	Ser	Gly	Gly	Phe	Ser	Val	Ser	Arg	
				110					115						120
Ala	Thr	Thr	Pro	Gln	Leu	Thr	Asp	Val	Ser	Glu	Asp	Leu	Asp	Gln	
				125					130						135
Gln	Ser	Leu	Pro	Ser	Glu	Pro	Glu	Glu	Thr	Leu	Ser	Arg	Asp	Leu	
				140					145						150
Gly	Glu	Gly	Glu	Glu	Gly	Glu	Leu	Ala	Pro	Pro	Glu	Asp	Leu	Leu	
				155					160						165
Gly	Arg	Pro	Gln	Ala	Leu	Ser	Arg	Gln	Ala	Leu	Asp	Ser	Glu	Glu	
				170					175						180
Glu	Glu	Glu	Asp	Val	Ala	Ala	Lys	Glu	Thr	Leu	Leu	Arg	Leu	Ser	
				185					190						195
Ser	Pro	Leu	His	Phe	Val	Asn	Thr	His	Phe	Asn	Gly	Ala	Gly	Ser	
				200					205						210
Pro	Gln	Asp	Gly	Val	Lys	Cys	Ser	Pro	Gly	Gly	Pro	Val	Glu	Thr	
				215					220						225
Leu	Ser	Pro	Glu	Thr	Val	Ser	Gly	Gly	Leu	Thr	Ala	Leu	Pro	Gly	
				230					235						240
Thr	Leu	Ser	Pro	Pro	Leu	Cys	Leu	Val	Gly	Ser	Asp	Pro	Ala	Pro	
				245					250						255
Ser	Pro	Ser	Ile	Leu	Pro	Pro	Val	Pro	Gln	Asp	Ser	Pro	Gln	Pro	
				260					265						270
Leu	Pro	Ala	Pro	Glu	Glu	Glu	Glu	Ala	Leu	Thr	Thr	Glu	Asp	Phe	
				275					280						285
Glu	Leu	Leu	Asp	Gln	Gly	Glu	Leu	Glu	Gln	Leu	Asn	Ala	Glu	Leu	
				290					295						300
Gly	Leu	Glu	Pro	Glu	Thr	Pro	Pro	Lys	Pro	Pro	Asp	Ala	Pro	Pro	
				305					310						315
Leu	Gly	Pro	Asp	Ile	His	Ser	Leu	Val	Gln	Ser	Asp	Gln	Glu	Ala	

[illegible]

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 150 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25 :

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 217 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: UTRSNOT06  
(B) CLONE: 1638407

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26 :

Met Ala Pro Pro Ala Leu Gln Arg Gly Gln Arg Val Ala Ala Val  
5 10 15

PF-0459 US

Ala	Val	Gly	Ser	Gln	Ala	Val	Leu	Gln	Ile	Leu	Ser	Arg	Val	Ser	
				20					25					30	
Gly	Arg	Gln	Ala	Pro	Pro	Gln	Pro	Ser	Gly	Ser	Gly	Gly	Val	Gly	
				35					40					45	
Ala	Gly	Pro	Val	Val	Val	Pro	Asp	Gly	Gly	Gly	Glu	Gly	Pro	Gln	
				50					55					60	
Pro	His	Pro	Ser	Ser	Ser	Gln	Ser	Pro	Pro	Asp	Leu	Pro	Leu	Lys	
				65					70					75	
Ala	Gly	Asp	Thr	Val	Met	Gly	Lys	Gln	Ala	Gln	Arg	Asp	Ile	Arg	
				80					85					90	
Leu	Arg	Val	Arg	Ala	Glu	Tyr	Cys	Glu	His	Gly	Pro	Ala	Leu	Glu	
				95					100					105	
Gln	Gly	Val	Ala	Ser	Arg	Arg	Pro	Gln	Ala	Leu	Ala	Arg	Gln	Leu	
				110					115					120	
Asp	Val	Phe	Gly	Gln	Ala	Thr	Ala	Val	Leu	Arg	Ser	Arg	Asp	Leu	
				125					130					135	
Gly	Ser	Val	Val	Cys	Asp	Ile	Lys	Phe	Ser	Glu	Leu	Ser	Tyr	Leu	
				140					145					150	
Asp	Ala	Phe	Trp	Gly	Asp	Tyr	Leu	Ser	Gly	Ala	Leu	Leu	Gln	Ala	
				155					160					165	
Leu	Arg	Gly	Val	Phe	Leu	Thr	Glu	Ala	Leu	Arg	Glu	Ala	Val	Gly	
				170					175					180	
Arg	Glu	Ala	Val	Arg	Leu	Leu	Val	Ser	Val	Asp	Glu	Ala	Asp	Tyr	
				185					190					195	
Glu	Ala	Gly	Arg	Arg	Arg	Leu	Leu	Leu	Met	Ala	Glu	Glu	Gly	Gly	
				200					205					210	
Arg	Arg	Pro	Thr	Glu	Ala	Ser									
				215											

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 504 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PROSTUT08
- (B) CLONE: 1653112

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27 :

Met	Ser	Gln	Pro	Arg	Thr	Pro	Glu	Gln	Ala	Leu	Asp	Thr	Pro	Gly	
				5					10					15	
Asp	Cys	Pro	Pro	Gly	Arg	Arg	Asp	Glu	Asp	Ala	Gly	Glu	Gly	Ile	
				20					25					30	
Gln	Cys	Ser	Gln	Arg	Met	Leu	Ser	Phe	Ser	Asp	Ala	Leu	Leu	Ser	
				35					40					45	
Ile	Ile	Ala	Thr	Val	Met	Ile	Leu	Pro	Val	Thr	His	Thr	Glu	Ile	
				50					55					60	
Ser	Pro	Glu	Gln	Gln	Phe	Asp	Arg	Ser	Val	Gln	Arg	Leu	Leu	Ala	
				65					70					75	
Thr	Arg	Ile	Ala	Val	Tyr	Leu	Met	Thr	Phe	Leu	Ile	Val	Thr	Val	
				80					85					90	
Ala	Trp	Ala	Ala	His	Thr	Arg	Leu	Phe	Gln	Val	Val	Gly	Lys	Thr	
				95					100					105	
Asp	Asp	Thr	Leu	Ala	Leu	Leu	Asn	Leu	Ala	Cys	Met	Met	Thr	Ile	

Thr Phe Leu Pro	110	Tyr Thr Phe Ser Leu	115	Met Val Thr Phe Pro	120
	125		130		135
Val Pro Leu Gly	140	Ile Phe Leu Phe Cys	145	Val Cys Val Ile Ala	150
	155		160		165
Gly Val Val Gln	170	Ala Leu Ile Val Gly	175	Tyr Ala Phe His Phe	180
	185		190		195
His Leu Leu Ser	200	Pro Gln Ile Gln Arg	205	Ser Ala His Arg Ala	210
	215		220		225
Tyr Arg Arg His	230	Val Leu Gly Ile Val	235	Leu Gln Gly Pro Ala	240
	245		250		255
Cys Phe Ala Ala	260	Ala Ile Phe Ser Leu	265	Phe Phe Val Pro Leu	270
	275		280		285
Tyr Leu Leu Met	290	Val Thr Val Ile Leu	295	Leu Pro Tyr Val Ser	300
	305		310		315
Val Thr Gly Trp	320	Cys Arg Asp Arg Leu	325	Leu Gly His Arg Glu	330
	335		340		345
Ser Ala His Pro	350	Val Glu Val Phe Ser	355	Phe Asp Leu His Glu	360
	365		370		375
Leu Ser Lys Glu	380	Arg Val Glu Ala Phe	385	Ser Asp Gly Val Tyr	390
	395		400		405
Ile Val Ala Thr	410	Leu Leu Ile Leu Asp	415	Ile Cys Glu Asp Asn	420
	425		430		435
Pro Asp Pro Lys	440	Asp Val Lys Glu Arg	445	Phe Ser Gly Ser Leu	450
	455		460		465
Ala Ala Leu Ser	470	Ala Thr Gly Pro Arg	475	Phe Leu Ala Tyr Phe	480
	485		490		495
Ser Phe Ala Thr	500	Val Gly Leu Leu Trp		Phe Ala His His Ser	
Phe Leu His Val		Arg Lys Ala Thr Arg		Ala Met Gly Leu Leu	
Thr Leu Ser Leu		Ala Phe Val Gly Gly		Leu Pro Leu Ala Tyr	
Gln Thr Ser Ala		Phe Ala Arg Gln Pro		Arg Asp Glu Leu Glu	
Val Arg Val Ser		Cys Thr Ile Ile Phe		Leu Ala Ser Ile Phe	
Leu Ala Met Trp		Thr Thr Ala Leu Leu		His Gln Ala Glu Thr	
Gln Pro Ser Val		Trp Phe Gly Gly Arg		Glu His Val Leu Met	
Ala Lys Leu Ala		Leu Tyr Pro Cys Ala		Ser Leu Leu Ala Phe	
Ser Thr Cys Leu		Leu Ser Arg Phe Ser		Val Gly Ile Phe His	
Met Gln Ile Ala		Val Pro Cys Ala Phe		Leu Leu Leu Arg Leu	
Val Gly Leu Ala		Leu Ala Thr Leu Arg		Val Leu Arg Gly Leu	
Arg Pro Glu His		Pro Pro Pro Ala Pro		Thr Gly Gln Asp Asp	
Gln Ser Gln Leu		Leu Pro Ala Pro Cys			

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 320 amino acids

PF-0459 US

(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: BRSTNOT09  
(B) CLONE: 1664634

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28 :

Met	Ala	Ala	Arg	Leu	Asp	Gly	Gly	Phe	Ala	Ala	Val	Ser	Arg	Ala	
				5					10					15	
Phe	His	Glu	Ile	Arg	Ala	Arg	Asn	Pro	Ala	Phe	Gln	Pro	Gln	Thr	
				20					25					30	
Leu	Met	Asp	Phe	Gly	Ser	Gly	Thr	Gly	Ser	Val	Thr	Trp	Ala	Ala	
				35					40					45	
His	Ser	Ile	Trp	Gly	Gln	Ser	Leu	Arg	Glu	Tyr	Met	Cys	Val	Asp	
				50					55					60	
Arg	Ser	Ala	Ala	Met	Leu	Val	Leu	Ala	Glu	Lys	Leu	Leu	Thr	Gly	
				65					70					75	
Gly	Ser	Glu	Ser	Gly	Glu	Pro	Tyr	Ile	Pro	Gly	Val	Phe	Phe	Arg	
				80					85					90	
Gln	Phe	Leu	Pro	Val	Ser	Pro	Lys	Val	Gln	Phe	Asp	Val	Val	Val	
				95					100					105	
Ser	Ala	Phe	Ser	Leu	Ser	Asp	Gln	Leu	Leu	Thr	Phe	Ile	Leu	Ser	
				110					115					120	
Cys	Asn	Ser	Ser	Leu	Leu	His	Ile	Phe	Pro	Phe	Cys	Glu	Gln	Val	
				125					130					135	
Leu	Val	Glu	Asn	Gly	Thr	Lys	Ala	Gly	His	Ser	Leu	Leu	Met	Asp	
				140					145					150	
Ala	Arg	Asp	Leu	Val	Leu	Lys	Gly	Lys	Glu	Lys	Ser	Pro	Leu	Asp	
				155					160					165	
Pro	Arg	Pro	Gly	Phe	Val	Phe	Ala	Pro	Cys	Pro	His	Glu	Leu	Pro	
				170					175					180	
Cys	Pro	Gln	Leu	Thr	Asn	Leu	Ala	Cys	Ser	Phe	Ser	Gln	Ala	Tyr	
				185					190					195	
His	Pro	Ile	Pro	Phe	Ser	Trp	Asn	Lys	Lys	Pro	Lys	Glu	Glu	Lys	
				200					205					210	
Phe	Ser	Met	Val	Ile	Leu	Ala	Arg	Gly	Ser	Pro	Glu	Glu	Ala	His	
				215					220					225	
Arg	Trp	Pro	Arg	Ile	Thr	Gln	Pro	Val	Leu	Lys	Arg	Pro	Arg	His	
				230					235					240	
Val	His	Cys	His	Leu	Cys	Cys	Pro	Asp	Gly	His	Met	Gln	His	Ala	
				245					250					255	
Val	Leu	Thr	Ala	Arg	Arg	His	Gly	Arg	Tyr	Gly	Gly	Cys	Asp	Gln	
				260					265					270	
Asn	Gln	Trp	Asp	Val	Ala	Gly	Ser	Cys	Ser	Pro	Arg	Gln	His	Leu	
				275					280					285	
Phe	Pro	Gln	Gly	Phe	Val	Ser	Leu	Cys	Pro	Cys	Gln	Leu	Leu	Gly	
				290					295					300	
Arg	Ser	Phe	Thr	Cys	Ala	Tyr	Ser	Val	Cys	Val	Ser	Ser	Ile	Tyr	
				305					310					315	
Gly	Ser	Gly	Ser	Leu											
				320											

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

RAW SEQUENCE LISTING  
PATENT APPLICATION US/09/002,485

DATE: 02/07/98  
TIME: 14:21:12

INPUT SET: S23256.raw

This Raw Listing contains the General Information Section and up to the first 5 pages.

SEQUENCE LISTING

(1) General Information:

(i) APPLICANT: Lal, Preeti  
Hillman, Jennifer L.  
Corley, Neil C.  
Guegler, Karl J.  
Baugh, Mariah  
Sather, Susan  
Shah, Purvi

(ii) TITLE OF INVENTION: HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS

(iii) NUMBER OF SEQUENCES: 154

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
(B) STREET: 3174 PORTER DRIVE  
(C) CITY: PALO ALTO  
(D) STATE: CALIFORNIA  
(E) COUNTRY: USA  
(F) ZIP: 94304

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: TO BE ASSIGNED  
(B) FILING DATE: HERewith  
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: BILLINGS, LUCY J.  
(B) REGISTRATION NUMBER: 36,749  
(C) REFERENCE/DOCKET NUMBER: PF-0459 US

ENTERED

467227 23120060

-->

236

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

# RAW SEQUENCE LISTING PATENT APPLICATION US/09/002,485

DATE: 02/07/98  
TIME: 14:21:16

INPUT SET: S23256.raw

47 (ix) TELECOMMUNICATION INFORMATION:  
48 (A) TELEPHONE: (650) 855-0555  
49 (B) TELEFAX: (650) 845-4166  
50  
51  
52  
53 (2) INFORMATION FOR SEQ ID NO: 1:  
54  
55 (i) SEQUENCE CHARACTERISTICS:  
56 (A) LENGTH: 348 amino acids  
57 (B) TYPE: amino acid  
58 (C) STRANDEDNESS: single  
59 (D) TOPOLOGY: linear  
60  
61 (vii) IMMEDIATE SOURCE:  
62 (A) LIBRARY: HEARNOT01  
63 (B) CLONE: 305841  
64  
65 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1 :  
66  
67 Met Ala Ala Thr Leu Gly Pro Leu Gly Ser Trp Gln Gln Trp Arg  
68 5 10 15  
69 Arg Cys Leu Ser Ala Arg Asp Gly Ser Arg Met Leu Leu Leu Leu  
70 20 25 30  
71 Leu Leu Leu Gly Ser Gly Gln Gly Pro Gln Gln Val Gly Ala Gly  
72 35 40 45  
73 Gln Thr Phe Glu Tyr Leu Lys Arg Glu His Ser Leu Ser Lys Pro  
74 50 55 60  
75 Tyr Gln Gly Val Gly Thr Gly Ser Ser Ser Leu Trp Asn Leu Met  
76 65 70 75  
77 Gly Asn Ala Met Val Met Thr Gln Tyr Ile Arg Leu Thr Pro Asp  
78 80 85 90  
79 Met Gln Ser Lys Gln Gly Ala Leu Trp Asn Arg Val Pro Cys Phe  
80 95 100 105  
81 Leu Arg Asp Trp Glu Leu Gln Val His Phe Lys Ile His Gly Gln  
82 110 115 120  
83 Gly Lys Lys Asn Leu His Gly Asp Gly Leu Ala Ile Trp Tyr Thr  
84 125 130 135  
85 Lys Asp Arg Met Gln Pro Gly Pro Val Phe Gly Asn Met Asp Lys  
86 140 145 150  
87 Phe Val Gly Leu Gly Val Phe Val Asp Thr Tyr Pro Asn Glu Glu  
88 155 160 165  
89 Lys Gln Gln Glu Arg Val Phe Pro Tyr Ile Ser Ala Met Val Asn  
90 170 175 180  
91 Asn Gly Ser Leu Ser Tyr Asp His Glu Arg Asp Gly Arg Pro Thr  
92 185 190 195  
93 Glu Leu Gly Gly Cys Thr Ala Ile Val Arg Asn Leu His Tyr Asp  
94 200 205 210  
95 Thr Phe Leu Val Ile Arg Tyr Val Lys Arg His Leu Thr Ile Met  
96 215 220 225  
97 Met Asp Ile Asp Gly Lys His Glu Trp Arg Asp Cys Ile Glu Val  
98 230 235 240  
99 Pro Gly Val Arg Leu Pro Arg Gly Tyr Tyr Phe Gly Thr Ser Ser



# RAW SEQUENCE LISTING PATENT APPLICATION US/09/002,485

DATE: 02/07/98  
TIME: 14:21:19

INPUT SET: S23256.raw

100		245		250		255
101	Ile Thr Gly Asp	Leu Ser Asp Asn His	Asp Val Ile Ser Leu Lys			
102		260		265		270
103	Leu Phe Glu Leu	Thr Val Glu Arg Thr	Pro Glu Glu Glu Lys Leu			
104		275		280		285
105	His Arg Asp Val	Phe Leu Pro Ser Val	Asp Asn Met Lys Leu Pro			
106		290		295		300
107	Glu Met Thr Ala	Pro Leu Pro Pro Leu	Ser Gly Leu Ala Leu Phe			
108		305		310		315
109	Leu Ile Val Phe	Phe Ser Leu Val Phe	Ser Val Phe Ala Ile Val			
110		320		325		330
111	Ile Gly Ile Ile	Leu Tyr Asn Lys Trp	Gln Glu Gln Ser Arg Lys			
112		335		340		345
113	Arg Phe Tyr					
114						
115						
116						
117						

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 194 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: EOSIHET02
- (B) CLONE: 322866

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2 :

132	Met Gly Met Ser Ser Leu Lys Leu Leu Lys Tyr Val Leu Phe Phe	
133		5 10 15
134	Phe Asn Leu Leu Phe Trp Ile Cys Gly Cys Cys Ile Leu Gly Phe	
135		20 25 30
136	Gly Ile Tyr Leu Leu Ile His Asn Asn Phe Gly Val Leu Phe His	
137		35 40 45
138	Asn Leu Pro Ser Leu Thr Leu Gly Asn Val Phe Val Ile Val Gly	
139		50 55 60
140	Ser Ile Ile Met Val Val Ala Phe Leu Gly Cys Met Gly Ser Ile	
141		65 70 75
142	Lys Glu Asn Lys Cys Leu Leu Met Ser Phe Phe Ile Leu Leu Leu	
143		80 85 90
144	Ile Ile Leu Leu Ala Glu Val Thr Leu Ala Ile Leu Leu Phe Val	
145		95 100 105
146	Tyr Glu Gln Lys Leu Asn Glu Tyr Val Ala Lys Gly Leu Thr Asp	
147		110 115 120
148	Ser Ile His Arg Tyr His Ser Asp Asn Ser Thr Lys Ala Ala Trp	
149		125 130 135
150	Asp Ser Ile Gln Ser Phe Leu Gln Cys Cys Gly Ile Asn Gly Thr	
151		140 145 150
152	Ser Asp Leu Asp Ser Gly Ser Pro Ala Ser Cys Pro Ser Asp Arg	

# RAW SEQUENCE LISTING PATENT APPLICATION US/09/002,485

DATE: 02/07/98  
TIME: 14:21:23

INPUT SET: S23256.raw

153 155 160 165  
154 Lys Val Glu Gly Cys Tyr Ala Lys Glu Asp Phe Gly Phe Ile Gln  
155 170 175 180  
156 Phe Pro Val Tyr Arg Asn His His His Leu Cys Met Cys Asp  
157 185 190  
158  
159  
160  
161

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 342 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BEPINOT01
- (B) CLONE: 546656

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3 :

176 Met Ser Leu His Gly Lys Arg Lys Glu Ile Tyr Lys Tyr Glu Ala  
177 5 10 15  
178 Pro Trp Thr Val Tyr Ala Met Asn Trp Ser Val Arg Pro Asp Lys  
179 20 25 30  
180 Arg Phe Arg Leu Ala Leu Gly Ser Phe Val Glu Glu Tyr Asn Asn  
181 35 40 45  
182 Lys Val Gln Leu Val Gly Leu Asp Glu Glu Ser Ser Glu Phe Ile  
183 50 55 60  
184 Cys Arg Asn Thr Phe Asp His Pro Tyr Pro Thr Thr Lys Leu Met  
185 65 70 75  
186 Trp Ile Pro Asp Thr Lys Gly Val Tyr Pro Asp Leu Leu Ala Thr  
187 80 85 90  
188 Ser Gly Asp Tyr Leu Arg Val Trp Arg Val Gly Glu Thr Glu Thr  
189 95 100 105  
190 Arg Leu Glu Cys Leu Leu Asn Asn Asn Lys Asn Ser Asp Phe Cys  
191 110 115 120  
192 Ala Pro Leu Thr Ser Phe Asp Trp Asn Glu Val Asp Pro Tyr Leu  
193 125 130 135  
194 Leu Gly Thr Ser Ser Ile Asp Thr Thr Cys Thr Ile Trp Gly Leu  
195 140 145 150  
196 Glu Thr Gly Gln Val Leu Gly Arg Val Asn Leu Val Ser Gly His  
197 155 160 165  
198 Val Lys Thr Gln Leu Ile Ala His Asp Lys Glu Val Tyr Asp Ile  
199 170 175 180  
200 Ala Phe Ser Arg Ala Gly Gly Gly Arg Asp Met Phe Ala Ser Val  
201 185 190 195  
202 Gly Ala Asp Gly Ser Val Arg Met Phe Asp Leu Arg His Leu Glu  
203 200 205 210  
204 His Ser Thr Ile Ile Tyr Glu Asp Pro Gln His His Pro Leu Leu  
205 215 220 225

153 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225

# RAW SEQUENCE LISTING PATENT APPLICATION US/09/002,485

DATE: 02/07/98  
TIME: 14:21:27

INPUT SET: S23256.raw

```

206 Arg Leu Cys Trp Asn Lys Gln Asp Pro Asn Tyr Leu Ala Thr Met
207                               230                               235                               240
208 Ala Met Asp Gly Met Glu Val Val Ile Leu Asp Val Arg Val Pro
209                               245                               250                               255
210 Cys Thr Pro Val Ala Arg Leu Asn Asn His Arg Ala Cys Val Asn
211                               260                               265                               270
212 Gly Ile Ala Trp Ala Pro His Ser Ser Cys His Ile Cys Thr Ala
213                               275                               280                               285
214 Ala Asp Asp His Gln Ala Leu Ile Trp Asp Ile Gln Gln Met Pro
215                               290                               295                               300
216 Arg Ala Ile Glu Asp Pro Ile Leu Ala Tyr Thr Ala Glu Gly Glu
217                               305                               310                               315
218 Ile Asn Asn Val Gln Trp Ala Ser Thr Gln Pro Asp Trp Ile Ala
219                               320                               325                               330
220 Ile Cys Tyr Asn Asn Cys Leu Glu Ile Leu Arg Val
221                               335                               340
222
223
224
225

```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 656 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: SYNORAT03
- (B) CLONE: 693453

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4 :

```

240 Met Glu Glu Leu Asp Gly Glu Pro Thr Val Thr Leu Ile Pro Gly
241                               5                               10                               15
242 Val Asn Ser Lys Lys Asn Gln Met Tyr Phe Asp Trp Gly Pro Gly
243                               20                               25                               30
244 Glu Met Leu Val Cys Glu Thr Ser Phe Asn Lys Lys Glu Lys Ser
245                               35                               40                               45
246 Glu Met Val Pro Ser Cys Pro Phe Ile Tyr Ile Ile Arg Lys Asp
247                               50                               55                               60
248 Val Asp Val Tyr Ser Gln Ile Leu Arg Lys Leu Phe Asn Glu Ser
249                               65                               70                               75
250 His Gly Ile Phe Leu Gly Leu Gln Arg Ile Asp Glu Glu Leu Thr
251                               80                               85                               90
252 Gly Lys Ser Arg Lys Ser Gln Leu Val Arg Val Ser Lys Asn Tyr
253                               95                               100                              105
254 Arg Ser Val Ile Arg Ala Cys Met Glu Glu Met His Gln Val Ala
255                               110                              115                              120
256 Ile Ala Ala Lys Asp Pro Ala Asn Gly Arg Gln Phe Ser Ser Gln
257                               125                              130                              135
258 Val Ser Ile Leu Ser Ala Met Glu Leu Ile Trp Asn Leu Cys Glu

```

**INPUT SET: S23256.raw**

Line	Error	Original Text
36	Wrong application Serial Number	(A) APPLICATION NUMBER: TO BE ASSIGNED

PF-0459 US

(A) LENGTH: 117 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: PROSTUT10  
(B) CLONE: 1690990

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29 :

Met	Asp	Asn	Lys	Gly	Ile	Tyr	Pro	Gly	Ala	Val	Phe	Tyr	His	Asp	
				5					10					15	
Ser	Phe	Thr	Glu	Ser	Arg	Val	Val	Leu	Leu	Arg	Ile	Arg	Thr	Leu	
			20						25					30	
Val	Pro	Tyr	Ser	Pro	Pro	Asp	Cys	Pro	Thr	Thr	Thr	Thr	Ala	Tyr	
			35						40					45	
Ser	Pro	Phe	Pro	Asn	His	Gly	Gln	Gln	Ile	Glu	Leu	Leu	Thr	Glu	
			50						55					60	
Val	Ser	Phe	Arg	Trp	Ile	Ser	Gln	Pro	Phe	Pro	His	Arg	Pro	His	
			65						70					75	
Arg	Glu	Thr	Val	Thr	Asp	Cys	Tyr	Ser	Pro	Asn	Thr	Gln	Val	Lys	
			80						85					90	
Ser	Asn	Ala	Gly	Arg	Asn	Asn	Ser	Lys	Ser	Phe	Asn	Phe	Leu	Ile	
			95						100					105	
Leu	Leu	Leu	Lys	Ile	Leu	Thr	Glu	Ala	Ser	Arg	Phe				
			110						115						

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 298 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: DUODNOT02  
(B) CLONE: 1704050

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30 :

Met	Ala	Arg	Arg	Ser	Arg	His	Arg	Leu	Leu	Leu	Leu	Leu	Leu	Arg	
				5					10					15	
Tyr	Leu	Val	Val	Ala	Leu	Gly	Tyr	His	Lys	Ala	Tyr	Gly	Phe	Ser	
			20						25					30	
Ala	Pro	Lys	Asp	Gln	Gln	Val	Val	Thr	Ala	Val	Glu	Tyr	Gln	Glu	
			35						40					45	
Ala	Ile	Leu	Ala	Cys	Lys	Thr	Pro	Lys	Lys	Thr	Val	Ser	Ser	Arg	
			50						55					60	
Leu	Glu	Trp	Lys	Lys	Leu	Gly	Arg	Ser	Val	Ser	Phe	Val	Tyr	Tyr	
			65						70					75	
Gln	Gln	Thr	Leu	Gln	Gly	Asp	Phe	Lys	Asn	Arg	Ala	Glu	Met	Ile	
			80						85					90	
Asp	Phe	Asn	Ile	Arg	Ile	Lys	Asn	Val	Thr	Arg	Ser	Asp	Ala	Gly	
			95						100					105	
Lys	Tyr	Arg	Cys	Glu	Val	Ser	Ala	Pro	Ser	Glu	Gln	Gly	Gln	Asn	
			110						115					120	

PF-0459 US

Leu	Glu	Glu	Asp	Thr	Val	Thr	Leu	Glu	Val	Leu	Val	Ala	Pro	Ala		
				125					130							135
Val	Pro	Ser	Cys	Glu	Val	Pro	Ser	Ser	Ala	Leu	Ser	Gly	Thr	Val		
				140					145							150
Val	Glu	Leu	Arg	Cys	Gln	Asp	Lys	Glu	Gly	Asn	Pro	Ala	Pro	Glu		
				155					160							165
Tyr	Thr	Trp	Phe	Lys	Asp	Gly	Ile	Arg	Leu	Leu	Glu	Asn	Pro	Arg		
				170					175							180
Leu	Gly	Ser	Gln	Ser	Thr	Asn	Ser	Ser	Tyr	Thr	Met	Asn	Thr	Lys		
				185					190							195
Thr	Gly	Thr	Leu	Gln	Phe	Asn	Thr	Val	Ser	Lys	Leu	Asp	Thr	Gly		
				200					205							210
Glu	Tyr	Ser	Cys	Glu	Ala	Arg	Asn	Ser	Val	Gly	Tyr	Arg	Arg	Cys		
				215					220							225
Pro	Gly	Lys	Arg	Met	Gln	Val	Asp	Asp	Leu	Asn	Ile	Ser	Gly	Ile		
				230					235							240
Ile	Ala	Ala	Val	Val	Val	Val	Ala	Leu	Val	Ile	Ser	Val	Cys	Gly		
				245					250							255
Leu	Gly	Val	Cys	Tyr	Ala	Gln	Arg	Lys	Gly	Tyr	Phe	Ser	Lys	Glu		
				260					265							270
Thr	Ser	Phe	Gln	Lys	Ser	Asn	Ser	Ser	Ser	Lys	Ala	Thr	Thr	Met		
				275					280							285
Ser	Glu	Asn	Asp	Phe	Lys	His	Thr	Lys	Ser	Phe	Ile	Ile				
				290					295							

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 118 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PROSNOT16
- (B) CLONE: 1711840

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31 :

Met	Gln	His	Arg	Gly	Phe	Leu	Leu	Leu	Thr	Leu	Leu	Ala	Leu	Leu		
				5					10							15
Ala	Leu	Thr	Ser	Ala	Val	Ala	Lys	Lys	Gln	Asp	Lys	Val	Lys	Lys		
				20					25							30
Gly	Gly	Pro	Gly	Ser	Glu	Cys	Ala	Glu	Trp	Ala	Trp	Gly	Pro	Cys		
				35					40							45
Thr	Pro	Ser	Ser	Lys	Gly	Phe	Ala	Ala	Val	Gly	Phe	Pro	Arg	Gly		
				50					55							60
Pro	Pro	Trp	Gly	Gly	Pro	Arg	Thr	Gln	Pro	Ala	Val	Leu	Val	Glu		
				65					70							75
Arg	Val	Ala	Pro	Gly	Lys	Leu	Glu	Arg	Lys	Glu	Phe	Trp	Ala	Pro		
				80					85							90
Gly	Leu	Trp	Lys	Val	Gly	Gln	Ile	Phe	Trp	Lys	Lys	Thr	Trp	Arg		
				95					100							105
Val	Cys	Arg	Ser	Val	Lys	Trp	Gly	Arg	Gly	Gln	Lys	Asn				
				110					115							

## (2) INFORMATION FOR SEQ ID NO: 32:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 248 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32 :

Met	Gln	Thr	Cys	Pro	Leu	Ala	Phe	Pro	Gly	His	Val	Ser	Gln	Ala	
				5					10					15	
Leu	Gly	Thr	Leu	Leu	Phe	Leu	Ala	Ala	Ser	Leu	Ser	Ala	Gln	Asn	
				20					25					30	
Glu	Gly	Trp	Asp	Ser	Pro	Ile	Cys	Thr	Glu	Gly	Val	Val	Ser	Val	
				35					40					45	
Ser	Trp	Gly	Glu	Asn	Thr	Val	Met	Ser	Cys	Asn	Ile	Ser	Asn	Ala	
				50					55					60	
Phe	Ser	His	Val	Asn	Ile	Lys	Leu	Arg	Ala	His	Gly	Gln	Glu	Ser	
				65					70					75	
Ala	Ile	Phe	Asn	Glu	Val	Ala	Pro	Gly	Tyr	Phe	Ser	Arg	Asp	Gly	
				80					85					90	
Trp	Gln	Leu	Gln	Val	Gln	Gly	Gly	Val	Ala	Gln	Leu	Val	Ile	Lys	
				95					100					105	
Gly	Ala	Arg	Asp	Ser	His	Ala	Gly	Leu	Tyr	Met	Trp	His	Leu	Val	
				110					115					120	
Gly	His	Gln	Arg	Asn	Asn	Arg	Gln	Val	Thr	Leu	Glu	Val	Ser	Gly	
				125					130					135	
Ala	Glu	Pro	Gln	Ser	Ala	Pro	Asp	Thr	Gly	Phe	Trp	Pro	Val	Pro	
				140					145					150	
Ala	Val	Val	Thr	Ala	Val	Phe	Ile	Leu	Leu	Val	Ala	Leu	Val	Met	
				155					160					165	
Phe	Ala	Trp	Tyr	Arg	Cys	Arg	Cys	Ser	Gln	Gln	Arg	Arg	Glu	Lys	
				170					175					180	
Lys	Phe	Phe	Leu	Leu	Glu	Pro	Gln	Met	Lys	Val	Ala	Ala	Leu	Arg	
				185					190					195	
Ala	Gly	Ala	Gln	Gln	Gly	Leu	Ser	Arg	Ala	Ser	Ala	Glu	Leu	Trp	
				200					205					210	
Thr	Pro	Asp	Ser	Glu	Pro	Thr	Pro	Arg	Pro	Leu	Ala	Leu	Val	Phe	
				215					220					225	
Lys	Pro	Ser	Pro	Leu	Gly	Ala	Leu	Glu	Leu	Leu	Ser	Pro	Gln	Pro	
				230					235					240	
Leu	Phe	Pro	Tyr	Ala	Ala	Asp	Pro								
				245											

## (2) INFORMATION FOR SEQ ID NO: 33:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 150 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: STOMTUT02

(B) CLONE: 1750632

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33 :

Met	Leu	Glu	Glu	Gly	Ser	Phe	Arg	Gly	Arg	Thr	Ala	Asp	Phe	Val	5	10	15
Phe	Met	Phe	Leu	Phe	Gly	Gly	Val	Leu	Met	Thr	Val	Ser	Phe	Pro	20	25	30
Gln	Ala	Leu	Glu	Pro	Arg	Ala	Arg	Ala	Pro	Arg	Arg	Pro	Ala	Cys	35	40	45
Val	Gly	Pro	Gly	Ala	Asn	Thr	Ala	Met	Pro	Glu	Arg	Asp	Thr	Val	50	55	60
Ala	Val	Ser	Ser	Leu	Ala	Pro	Phe	Leu	Pro	Trp	Ala	Leu	Met	Gly	65	70	75
Phe	Ser	Leu	Leu	Leu	Gly	Asn	Ser	Ile	Leu	Val	Asp	Leu	Leu	Gly	80	85	90
Ile	Ala	Val	Gly	His	Ile	Tyr	Tyr	Phe	Leu	Glu	Asp	Val	Phe	Pro	95	100	105
Asn	Gln	Pro	Gly	Gly	Lys	Arg	Leu	Leu	Gln	Thr	Pro	Gly	Phe	Leu	110	115	120
Lys	Leu	Leu	Leu	Asp	Ala	Pro	Ala	Glu	Asp	Pro	Asn	Tyr	Leu	Pro	125	130	135
Leu	Pro	Glu	Glu	Gln	Pro	Gly	Pro	His	Leu	Pro	Pro	Pro	Gln	Gln	140	145	150

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 431 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34 :

Met	Trp	Ala	Leu	Gly	Gln	Ala	Gly	Phe	Ala	Asn	Leu	Thr	Glu	Gly	5	10	15
Leu	Lys	Val	Trp	Leu	Gly	Ile	Met	Leu	Pro	Val	Leu	Gly	Ile	Lys	20	25	30
Ser	Leu	Ser	Pro	Phe	Ala	Ile	Thr	Tyr	Leu	Asp	Arg	Leu	Leu	Leu	35	40	45
Met	His	Pro	Asn	Leu	Thr	Lys	Gly	Phe	Gly	Met	Ile	Gly	Pro	Lys	50	55	60
Asp	Phe	Phe	Pro	Leu	Leu	Asp	Phe	Ala	Tyr	Met	Pro	Asn	Asn	Ser	65	70	75
Leu	Thr	Pro	Ser	Leu	Gln	Glu	Gln	Leu	Cys	Gln	Leu	Tyr	Pro	Arg	80	85	90
Leu	Lys	Met	Leu	Ala	Phe	Gly	Ala	Lys	Pro	Asp	Ser	Thr	Leu	His	95	100	105
Thr	Tyr	Phe	Pro	Ser	Phe	Leu	Ser	Arg	Ala	Thr	Pro	Ser	Cys	Pro	110	115	120



Pro	Glu	Met	Lys	Lys	Glu	Leu	Leu	Ser	Ser	Leu	Thr	Glu	Cys	Leu	
				125					130					135	
Thr	Val	Asp	Pro	Leu	Ser	Ala	Ser	Val	Trp	Arg	Gln	Leu	Tyr	Pro	
				140					145					150	
Lys	His	Leu	Ser	Gln	Ser	Ser	Leu	Leu	Leu	Glu	His	Leu	Leu	Ser	
				155					160					165	
Ser	Trp	Glu	Gln	Ile	Pro	Lys	Lys	Val	Gln	Lys	Ser	Leu	Gln	Glu	
				170					175					180	
Thr	Ile	Gln	Ser	Leu	Lys	Leu	Thr	Asn	Gln	Glu	Leu	Leu	Arg	Lys	
				185					190					195	
Gly	Ser	Ser	Asn	Asn	Gln	Asp	Val	Val	Thr	Cys	Asp	Met	Ala	Cys	
				200					205					210	
Lys	Gly	Leu	Leu	Gln	Gln	Val	Gln	Gly	Pro	Arg	Leu	Pro	Trp	Thr	
				215					220					225	
Arg	Leu	Leu	Leu	Leu	Leu	Leu	Val	Phe	Ala	Val	Gly	Phe	Leu	Cys	
				230					235					240	
His	Asp	Leu	Arg	Ser	His	Ser	Ser	Phe	Gln	Ala	Ser	Leu	Thr	Gly	
				245					250					255	
Arg	Leu	Leu	Arg	Ser	Ser	Gly	Phe	Leu	Pro	Ala	Ser	Gln	Gln	Ala	
				260					265					270	
Cys	Ala	Lys	Leu	Tyr	Ser	Tyr	Ser	Leu	Gln	Gly	Tyr	Ser	Trp	Leu	
				275					280					285	
Gly	Glu	Thr	Leu	Pro	Leu	Trp	Gly	Ser	His	Leu	Leu	Thr	Val	Val	
				290					295					300	
Arg	Pro	Ser	Leu	Gln	Leu	Ala	Trp	Ala	His	Thr	Asn	Ala	Thr	Val	
				305					310					315	
Ser	Phe	Leu	Ser	Ala	His	Cys	Ala	Ser	His	Leu	Ala	Trp	Phe	Gly	
				320					325					330	
Asp	Ser	Leu	Thr	Ser	Leu	Ser	Gln	Arg	Leu	Gln	Ile	Gln	Leu	Pro	
				335					340					345	
Asp	Ser	Val	Asn	Gln	Leu	Leu	Arg	Tyr	Leu	Arg	Glu	Leu	Pro	Leu	
				350					355					360	
Leu	Phe	His	Gln	Asn	Val	Leu	Leu	Pro	Leu	Trp	His	Leu	Leu	Leu	
				365					370					375	
Glu	Ala	Leu	Ala	Trp	Ala	Gln	Glu	His	Cys	His	Glu	Ala	Cys	Arg	
				380					385					390	
Gly	Glu	Val	Thr	Trp	Asp	Cys	Met	Lys	Thr	Gln	Leu	Ser	Glu	Ala	
				395					400					405	
Val	His	Trp	Thr	Trp	Leu	Cys	Leu	Gln	Asp	Ile	Thr	Val	Ala	Phe	
				410					415					420	
Leu	Asp	Trp	Ala	Leu	Ala	Leu	Ile	Ser	Gln	Gln					
				425					430						

(2) INFORMATION FOR SEQ ID NO: 35:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 278 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PROSNOT20
- (B) CLONE: 1818761

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35 :

Met	Gln	Trp	Leu	Arg	Val	Arg	Glu	Ser	Pro	Gly	Glu	Ala	Thr	Gly			
				5					10					15			
His	Arg	Val	Thr	Met	Gly	Thr	Ala	Ala	Leu	Gly	Pro	Val	Trp	Ala			
				20					25					30			
Ala	Leu	Leu	Leu	Phe	Leu	Leu	Met	Cys	Glu	Ile	Pro	Met	Val	Glu			
				35					40					45			
Leu	Thr	Phe	Asp	Arg	Ala	Val	Ala	Ser	Gly	Cys	Gln	Arg	Cys	Cys			
				50					55					60			
Asp	Ser	Glu	Asp	Pro	Leu	Asp	Pro	Ala	His	Val	Ser	Ser	Ala	Ser			
				65					70					75			
Ser	Ser	Gly	Arg	Pro	His	Ala	Leu	Pro	Glu	Ile	Arg	Pro	Tyr	Ile			
				80					85					90			
Asn	Ile	Thr	Ile	Leu	Lys	Gly	Asp	Lys	Gly	Asp	Pro	Gly	Pro	Met			
				95					100					105			
Gly	Leu	Pro	Gly	Tyr	Met	Gly	Arg	Glu	Gly	Pro	Gln	Gly	Glu	Pro			
				110					115					120			
Gly	Pro	Gln	Gly	Ser	Lys	Gly	Asp	Lys	Gly	Glu	Met	Gly	Ser	Pro			
				125					130					135			
Gly	Ala	Pro	Cys	Gln	Lys	Arg	Phe	Phe	Ala	Phe	Ser	Val	Gly	Arg			
				140					145					150			
Lys	Thr	Ala	Leu	His	Ser	Gly	Glu	Asp	Phe	Gln	Thr	Leu	Leu	Phe			
				155					160					165			
Glu	Arg	Val	Phe	Val	Asn	Leu	Asp	Gly	Cys	Phe	Asp	Met	Ala	Thr			
				170					175					180			
Gly	Gln	Phe	Ala	Ala	Pro	Leu	Arg	Gly	Ile	Tyr	Phe	Phe	Ser	Leu			
				185					190					195			
Asn	Val	His	Ser	Trp	Asn	Tyr	Lys	Glu	Thr	Tyr	Val	His	Ile	Met			
				200					205					210			
His	Asn	Gln	Lys	Glu	Ala	Val	Ile	Leu	Tyr	Ala	Gln	Pro	Ser	Glu			
				215					220					225			
Arg	Ser	Ile	Met	Gln	Ser	Gln	Ser	Val	Met	Leu	Asp	Leu	Ala	Tyr			
				230					235					240			
Gly	Asp	Arg	Val	Trp	Val	Arg	Leu	Phe	Lys	Arg	Gln	Arg	Glu	Asn			
				245					250					255			
Ala	Ile	Tyr	Ser	Asn	Asp	Phe	Asp	Thr	Tyr	Ile	Thr	Phe	Ser	Gly			
				260					265					270			
His	Leu	Ile	Lys	Ala	Glu	Asp	Asp										
				275													

## (2) INFORMATION FOR SEQ ID NO: 36:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 286 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GBLATUT01
- (B) CLONE: 1824469

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36 :

Met	Glu	Glu	Lys	Arg	Arg	Arg	Ala	Arg	Val	Gln	Gly	Ala	Trp	Ala			
			5						10					15			
Ala	Pro	Val	Lys	Ser	Gln	Ala	Ile	Ala	Gln	Pro	Ala	Thr	Thr	Ala			

Lys	Ser	His	Leu	His	Gln	Lys	Pro	Gly	Gln	Thr	Trp	Lys	Asn	Lys	20	25	30
				35					40							45	
Glu	His	His	Leu	Ser	Asp	Arg	Glu	Phe	Val	Phe	Lys	Glu	Pro	Gln	50	55	60
Gln	Val	Val	Arg	Arg	Ala	Pro	Glu	Pro	Arg	Val	Ile	Asp	Arg	Glu	65	70	75
Gly	Val	Tyr	Glu	Ile	Ser	Leu	Ser	Pro	Thr	Gly	Val	Ser	Arg	Val	80	85	90
Cys	Leu	Tyr	Pro	Gly	Phe	Val	Asp	Val	Lys	Glu	Ala	Asp	Trp	Ile	95	100	105
Leu	Glu	Gln	Leu	Cys	Gln	Asp	Val	Pro	Trp	Lys	Gln	Arg	Thr	Gly	110	115	120
Ile	Arg	Glu	Asp	Ile	Thr	Tyr	Gln	Gln	Pro	Arg	Leu	Thr	Ala	Trp	125	130	135
Tyr	Gly	Glu	Leu	Pro	Tyr	Thr	Tyr	Ser	Arg	Ile	Thr	Met	Glu	Pro	140	145	150
Asn	Pro	His	Trp	His	Pro	Val	Leu	Arg	Thr	Leu	Lys	Asn	Arg	Ile	155	160	165
Glu	Glu	Asn	Thr	Gly	His	Thr	Phe	Asn	Ser	Leu	Leu	Cys	Asn	Leu	170	175	180
Tyr	Arg	Asn	Glu	Lys	Asp	Ser	Val	Asp	Trp	His	Ser	Asp	Asp	Glu	185	190	195
Pro	Ser	Leu	Gly	Arg	Cys	Pro	Ile	Ile	Ala	Ser	Leu	Ser	Phe	Gly	200	205	210
Ala	Thr	Arg	Thr	Phe	Glu	Met	Arg	Lys	Lys	Pro	Pro	Pro	Glu	Glu	215	220	225
Asn	Gly	Asp	Tyr	Thr	Tyr	Val	Glu	Arg	Val	Lys	Ile	Pro	Leu	Asp	230	235	240
His	Gly	Thr	Leu	Leu	Ile	Met	Glu	Gly	Ala	Thr	Gln	Ala	Asp	Trp	245	250	255
Gln	His	Arg	Val	Pro	Lys	Glu	Tyr	His	Ser	Arg	Glu	Pro	Arg	Val	260	265	270
Asn	Leu	Thr	Phe	Arg	Thr	Val	Tyr	Pro	Asp	Pro	Arg	Gly	Ala	Pro	275	280	285
Trp																	

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PROSNOT19
- (B) CLONE: 1864292

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37 :

Met	Lys	Met	Glu	Glu	Ala	Val	Gly	Lys	Val	Glu	Glu	Leu	Ile	Glu	5	10	15
Ser	Glu	Ala	Pro	Pro	Lys	Ala	Ser	Glu	Gln	Glu	Thr	Ala	Lys	Glu	20	25	30
Glu	Asp	Gly	Ser	Val	Glu	Leu	Glu	Ser	Gln	Val	Gln	Lys	Asp	Gly	35	40	45

PF-0459 US

Val	Ala	Asp	Ser	Thr	Val	Ile	Ser	Ser	Met	Pro	Cys	Leu	Leu	Met	50	55	60
Glu	Leu	Arg	Arg	Asp	Ser	Ser	Glu	Ser	Gln	Leu	Ala	Ser	Thr	Glu	65	70	75
Ser	Asp	Lys	Pro	Thr	Thr	Gly	Arg	Val	Tyr	Glu	Ser	Asp	Pro	Ser	80	85	90
Asn	His	Cys	Met	Leu	Ser	Pro	Ser	Ser	Ser	Gly	His	Leu	Ala	Asp	95	100	105
Ser	Asp	Thr	Leu	Ser	Ser	Ala	Glu	Glu	Asn	Glu	Pro	Ser	Gln	Ala	110	115	120
Glu	Thr	Ala	Val	Glu	Gly	Asp	Pro	Ser	Gly	Val	Ser	Gly	Ala	Thr	125	130	135
Val	Gly	Arg	Lys	Ser	Arg	Arg	Ser	Arg	Ser	Glu	Ser	Glu	Thr	Ser	140	145	150
Thr	Met	Ala	Ala	Lys	Lys	Asn	Arg	Gln	Ser	Ser	Asp	Lys	Gln	Asn	155	160	165
Gly	Arg	Val	Ala	Lys	Val	Lys	Gly	His	Arg	Ser	Gln	Lys	His	Lys	170	175	180
Glu	Arg	Ile	Arg	Leu	Leu	Arg	Gln	Lys	Arg	Glu	Ala	Ala	Ala	Arg	185	190	195
Lys	Lys	Tyr	Asn	Leu	Leu	Gln	Asp	Ser	Ser	Thr	Ser	Asp	Ser	Asp	200	205	210
Leu	Thr	Cys	Asp	Ser	Ser	Thr	Ser	Ser	Ser	Asp	Asp	Asp	Glu	Glu	215	220	225
Val	Ser	Gly	Ser	Ser	Lys	Thr	Ile	Thr	Ala	Glu	Ile	Pro	Asp	Gly	230	235	240
Pro	Pro	Val	Val	Ala	His	Tyr	Asp	Met	Ser	Asp	Thr	Asn	Ser	Asp	245	250	255
Pro	Glu	Val	Val	Asn	Val	Asp	Asn	Leu	Leu	Ala	Ala	Ala	Val	Val	260	265	270
Gln	Glu	His	Ser	Asn	Ser	Val	Gly	Gly	Gln	Asp	Thr	Gly	Ala	Thr	275	280	285
Trp	Arg	Thr	Ser	Gly	Leu	Leu	Glu	Glu	Leu	Asn	Ala	Glu	Ala	Gly	290	295	300
His	Leu	Asp	Pro	Gly	Phe	Leu	Ala	Ser	Asp	Lys	Thr	Ser	Ala	Gly	305	310	315
Asn	Ala	Pro	Leu	Asn	Glu	Glu	Ile	Asn	Ile	Ala	Ser	Ser	Asp	Ser	320	325	330
Glu	Val	Glu	Ile	Val	Gly	Val	Gln	Glu	His	Ala	Arg	Cys	Val	His	335	340	345
Pro	Arg	Gly	Gly	Val	Ile	Gln	Ser	Val	Ser	Ser	Trp	Lys	His	Gly	350	355	360
Ser	Gly	Thr	Gln	Tyr	Val	Ser	Thr	Arg	Gln	Thr	Gln	Ser	Trp	Thr	365	370	375
Ala	Val	Thr	Pro	Gln	Gln	Thr	Trp	Ala	Ser	Pro	Ala	Glu	Val	Val	380	385	390
Asp	Leu	Thr	Leu	Asp	Glu	Asp	Ser	Arg	Arg	Lys	Tyr	Leu	Leu		395	400	

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 405 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: THP1NOT01  
(B) CLONE: 1866437

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38 :

Met	Phe	Val	Gln	Glu	Glu	Lys	Ile	Phe	Ala	Gly	Lys	Val	Leu	Arg	5	10	15
Leu	His	Ile	Cys	Ala	Ser	Asp	Gly	Ala	Glu	Trp	Leu	Glu	Glu	Ala	20	25	30
Thr	Glu	Asp	Thr	Ser	Val	Glu	Lys	Leu	Lys	Glu	Arg	Cys	Leu	Lys	35	40	45
His	Cys	Ala	His	Gly	Ser	Leu	Glu	Asp	Pro	Lys	Ser	Ile	Thr	His	50	55	60
His	Lys	Leu	Ile	His	Ala	Ala	Ser	Glu	Arg	Val	Leu	Ser	Asp	Ala	65	70	75
Arg	Thr	Ile	Leu	Glu	Glu	Asn	Ile	Gln	Asp	Gln	Asp	Val	Leu	Leu	80	85	90
Leu	Lys	Lys	Lys	Arg	Ala	Pro	Ser	Pro	Leu	Pro	Lys	Met	Ala	Asp	95	100	105
Val	Ser	Ala	Glu	Glu	Lys	Lys	Lys	Gln	Asp	Gln	Lys	Ala	Pro	Asp	110	115	120
Lys	Glu	Ala	Ile	Leu	Arg	Ala	Thr	Ala	Asn	Leu	Pro	Ser	Tyr	Asn	125	130	135
Met	Asp	Arg	Ala	Ala	Val	Gln	Thr	Asn	Met	Arg	Asp	Phe	Gln	Thr	140	145	150
Glu	Leu	Arg	Lys	Ile	Leu	Val	Ser	Leu	Ile	Glu	Val	Ala	Gln	Lys	155	160	165
Leu	Leu	Ala	Leu	Asn	Pro	Asp	Ala	Val	Glu	Leu	Phe	Lys	Lys	Ala	170	175	180
Asn	Ala	Met	Leu	Asp	Glu	Asp	Glu	Asp	Glu	Arg	Val	Asp	Glu	Ala	185	190	195
Ala	Leu	Arg	Gln	Leu	Thr	Glu	Met	Gly	Phe	Pro	Glu	Asn	Arg	Ala	200	205	210
Thr	Lys	Ala	Leu	Gln	Leu	Asn	His	Met	Ser	Val	Pro	Gln	Ala	Met	215	220	225
Glu	Trp	Leu	Ile	Glu	His	Ala	Glu	Asp	Pro	Thr	Ile	Asp	Thr	Pro	230	235	240
Leu	Pro	Gly	Gln	Ala	Pro	Pro	Glu	Ala	Glu	Gly	Ala	Thr	Ala	Ala	245	250	255
Ala	Ser	Glu	Ala	Ala	Ala	Gly	Ala	Ser	Ala	Thr	Asp	Glu	Glu	Ala	260	265	270
Arg	Asp	Glu	Leu	Thr	Glu	Ile	Phe	Lys	Lys	Ile	Arg	Arg	Lys	Arg	275	280	285
Glu	Phe	Arg	Ala	Asp	Ala	Arg	Ala	Val	Ile	Ser	Leu	Met	Glu	Met	290	295	300
Gly	Phe	Asp	Glu	Lys	Glu	Val	Ile	Asp	Ala	Leu	Arg	Val	Asn	Asn	305	310	315
Asn	Gln	Gln	Asn	Ala	Ala	Cys	Glu	Trp	Leu	Leu	Gly	Asp	Arg	Lys	320	325	330
Pro	Ser	Pro	Glu	Glu	Leu	Asp	Lys	Gly	Ile	Asp	Pro	Asp	Ser	Pro	335	340	345
Leu	Phe	Gln	Ala	Ile	Leu	Asp	Asn	Pro	Val	Val	Gln	Leu	Gly	Leu	350	355	360
Thr	Asn	Pro	Lys	Thr	Leu	Leu	Ala	Phe	Glu	Asp	Met	Leu	Glu	Asn	365	370	375
Pro	Leu	Asn	Ser	Thr	Gln	Trp	Met	Asn	Asp	Pro	Glu	Thr	Gly	Pro	380	385	390
Val	Met	Leu	Gln	Ile	Ser	Arg	Ile	Phe	Gln	Thr	Leu	Asn	Arg	Thr	395	400	405

(2) INFORMATION FOR SEQ ID NO: 39:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 177 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: SKINBIT01  
 (B) CLONE: 1871375

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39 :

Met	Val	Met	His	Asn	Ser	Asp	Pro	Asn	Leu	His	Leu	Leu	Ala	Glu	5	10	15
Gly	Ala	Pro	Ile	Asp	Trp	Gly	Glu	Glu	Tyr	Ser	Asn	Ser	Gly	Gly	20	25	30
Gly	Gly	Ser	Pro	Ala	Pro	Ala	Pro	Arg	Ser	Gln	Pro	Pro	Ser	Arg	35	40	45
Lys	Ser	Asp	Gly	Ala	Pro	Ser	Arg	Trp	Ser	Leu	Trp	Ser	Arg	Met	50	55	60
Arg	Arg	Trp	Gly	Cys	Pro	Leu	Arg	Leu	Ala	Leu	Ser	His	His	His	65	70	75
Leu	Arg	Pro	Arg	Thr	Val	Ser	Leu	Arg	Ser	Glu	Ala	Cys	Trp	Pro	80	85	90
Lys	Val	Cys	Gly	Leu	Arg	Ala	Pro	His	Gln	Pro	Ala	Pro	Cys	Ser	95	100	105
Thr	Gly	Pro	Pro	Leu	Gly	Arg	Val	Pro	Ser	Leu	Arg	Pro	Pro	Pro	110	115	120
Arg	Pro	Pro	Arg	Arg	Leu	Pro	His	Pro	Ser	Ser	Ile	Ser	Cys	Leu	125	130	135
Glu	Arg	Leu	Trp	Thr	Leu	Gly	Pro	Pro	Ser	Pro	Ala	Thr	Arg	Arg	140	145	150
Leu	Glu	Ser	Arg	Cys	Pro	Ala	Pro	Ala	Ala	Thr	Pro	Pro	Ser	Thr	155	160	165
Pro	Pro	Pro	Arg	Xaa	Xaa	Phe	Lys	Gly	Cys	Lys	Asn				170	175	

(2) INFORMATION FOR SEQ ID NO: 40:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 197 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: LEUKNOT03  
 (B) CLONE: 1880830

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40 :

Met	Ile	Thr	Cys	Arg	Val	Cys	Gln	Ser	Leu	Ile	Asn	Val	Glu	Gly	5	10	15
Lys	Met	His	Gln	His	Val	Val	Lys	Cys	Gly	Val	Cys	Asn	Glu	Ala	20	25	30

PF-0459 US

Thr Pro Ile Lys Asn Ala Pro Pro Gly Lys Lys Tyr Val Arg Cys  
35 40 45  
Pro Cys Asn Cys Leu Leu Ile Cys Lys Val Thr Ser Gln Arg Ile  
50 55 60  
Ala Cys Pro Arg Pro Tyr Cys Lys Arg Ile Ile Asn Leu Gly Pro  
65 70 75  
Val His Pro Gly Pro Leu Ser Pro Glu Pro Gln Pro Met Gly Val  
80 85 90  
Arg Val Ile Cys Gly His Cys Lys Asn Thr Phe Leu Trp Thr Glu  
95 100 105  
Phe Thr Asp Arg Thr Leu Ala Arg Cys Pro His Cys Arg Lys Val  
110 115 120  
Ser Ser Ile Gly Arg Arg Tyr Pro Arg Lys Arg Cys Ile Cys Cys  
125 130 135  
Phe Leu Leu Gly Leu Leu Leu Ala Val Thr Ala Thr Gly Leu Ala  
140 145 150  
Phe Gly Thr Trp Lys His Ala Arg Arg Tyr Gly Gly Ile Tyr Ala  
155 160 165  
Ala Trp Ala Phe Val Ile Leu Leu Ala Val Leu Cys Leu Gly Arg  
170 175 180  
Ala Leu Tyr Trp Ala Cys Met Lys Val Ser His Pro Val Gln Asn  
185 190 195  
Phe Ser

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 302 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: OVARNOT07
- (B) CLONE: 1905325

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41 :

Met Leu Lys Asp Ile Ile Lys Glu Tyr Thr Asp Val Tyr Pro Glu  
5 10 15  
Ile Ile Glu Arg Ala Gly Tyr Ser Leu Glu Lys Val Phe Gly Ile  
20 25 30  
Gln Leu Lys Glu Ile Asp Lys Asn Asp His Leu Tyr Ile Leu Leu  
35 40 45  
Ser Thr Leu Glu Pro Thr Asp Ala Gly Ile Leu Gly Thr Thr Lys  
50 55 60  
Asp Ser Pro Lys Leu Gly Leu Leu Met Val Leu Leu Ser Ile Ile  
65 70 75  
Phe Met Asn Gly Asn Arg Ser Ser Glu Ala Val Ile Trp Glu Val  
80 85 90  
Leu Arg Lys Leu Gly Leu Arg Pro Gly Ile His His Ser Leu Phe  
95 100 105  
Gly Asp Val Lys Lys Leu Ile Thr Asp Glu Phe Val Lys Gln Lys  
110 115 120  
Tyr Leu Asp Tyr Ala Arg Val Pro Asn Ser Asn Pro Pro Glu Tyr  
125 130 135  
Glu Phe Phe Trp Gly Leu Arg Ser Tyr Tyr Glu Thr Ser Lys Met  
140 145 150  
Lys Val Leu Lys Phe Ala Cys Lys Val Gln Lys Lys Asp Pro Lys

Glu Trp Ala Ala	155	Tyr Arg Glu Ala	160	Glu Ala Asp Leu	165
Ala Ala Ala Glu	170	Ala Ala Glu Ala	175	Ala Arg Ala Glu	180
Arg Ala Arg Met	185	Gly Ile Gly Leu Gly	190	Ser Glu Asn Ala Ala	195
Pro Cys Asn Trp	200	Asp Glu Ala Asp Ile	205	Gly Pro Trp Ala Lys	210
Arg Ile Gln Ala	215	Gly Ala Glu Ala Lys	220	Ala Lys Ala Gln Glu	225
Gly Ser Ala Ser	230	Thr Gly Ala Ser Thr	235	Ser Thr Asn Asn Ser	240
Ser Ala Ser Ala	245	Ser Thr Ser Gly Gly	250	Phe Ser Ala Gly Ala	255
Leu Thr Ala Thr	260	Leu Thr Phe Gly Leu	265	Phe Ala Gly Leu Gly	270
Ala Gly Ala Ser	275	Thr Ser Gly Ser Ser	280	Gly Ala Cys Gly Phe	285
	290		295		300

Tyr Lys

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 164 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRSTTUT01
- (B) CLONE: 1919931

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42 :

Met Arg Thr Leu Glu Asn Gln Gly Phe Lys Ile Leu Pro Phe Leu	5	10	15
Gly Val Lys Glu Val Trp Gln Lys Gln Asn Lys Leu Ile Ser Arg	20	25	30
Phe Ile Thr Cys Gln Phe Phe Leu Tyr Asn Phe Leu Asp Ser Gly	35	40	45
Ser Ile Trp Val Gln Ala Asp Phe Pro Pro Ile Leu Gln Cys Gly	50	55	60
Cys Phe Leu Phe His Pro Trp Thr Leu Gln Glu Ile Ala Pro Cys	65	70	75
Phe Cys Leu Cys Ile Thr Glu Lys Gly Ser Met Lys Val Ala Gln	80	85	90
Val Arg Pro Phe His Cys Pro Pro Gly Ala Gly Phe Ala Leu Pro	95	100	105
Ile Leu Gly Leu Leu Gln Gly Leu Val Ile Leu His Ser Pro Leu	110	115	120
His Ile Ser Gln Val Ser Ala Gln Lys Ser Pro Phe Gly Gly Val	125	130	135
Ser Thr Cys His Cys Val Cys Lys Ser Ser Phe Ser Phe Phe Leu	140	145	150
Ala His Leu Thr Leu Val Met Ser Leu Ile Thr Thr Thr Ile	155	160	



	155		160		165
Glu Trp Ala Ala	Gln Tyr Arg Glu Ala	Met Glu Ala Asp Leu	Lys		
	170		175		180
Ala Ala Ala Glu	Ala Ala Ala Glu Ala	Lys Ala Arg Ala Glu	Ile		
	185		190		195
Arg Ala Arg Met	Gly Ile Gly Leu Gly	Ser Glu Asn Ala Ala	Gly		
	200		205		210
Pro Cys Asn Trp	Asp Glu Ala Asp Ile	Gly Pro Trp Ala Lys	Ala		
	215		220		225
Arg Ile Gln Ala	Gly Ala Glu Ala Lys	Ala Lys Ala Gln Glu	Ser		
	230		235		240
Gly Ser Ala Ser	Thr Gly Ala Ser Thr	Ser Thr Asn Asn Ser	Ala		
	245		250		255
Ser Ala Ser Ala	Ser Thr Ser Gly Gly	Phe Ser Ala Gly Ala	Ser		
	260		265		270
Leu Thr Ala Thr	Leu Thr Phe Gly Leu	Phe Ala Gly Leu Gly	Gly		
	275		280		285
Ala Gly Ala Ser	Thr Ser Gly Ser Ser	Gly Ala Cys Gly Phe	Ser		
	290		295		300
Tyr Lys					

## (2) INFORMATION FOR SEQ ID NO: 42:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 164 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRSTTUT01
- (B) CLONE: 1919931

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42 :

Met Arg Thr Leu Glu Asn Gln Gly Phe Lys Ile Leu Pro Phe Leu		
	5	10
Gly Val Lys Glu Val Trp Gln Lys Gln Asn Lys Leu Ile Ser Arg		
	20	25
Phe Ile Thr Cys Gln Phe Phe Leu Tyr Asn Phe Leu Asp Ser Gly		
	35	40
Ser Ile Trp Val Gln Ala Asp Phe Pro Pro Ile Leu Gln Cys Gly		
	50	55
Cys Phe Leu Phe His Pro Trp Thr Leu Gln Glu Ile Ala Pro Cys		
	65	70
Phe Cys Leu Cys Ile Thr Glu Lys Gly Ser Met Lys Val Ala Gln		
	80	85
Val Arg Pro Phe His Cys Pro Pro Gly Ala Gly Phe Ala Leu Pro		
	95	100
Ile Leu Gly Leu Leu Gln Gly Leu Val Ile Leu His Ser Pro Leu		
	110	115
His Ile Ser Gln Val Ser Ala Gln Lys Ser Pro Phe Gly Gly Val		
	125	130
Ser Thr Cys His Cys Val Cys Lys Ser Ser Phe Ser Phe Phe Leu		
	140	145
Ala His Leu Thr Leu Val Met Ser Leu Ile Thr Thr Thr Ile		
	155	160

## (2) INFORMATION FOR SEQ ID NO: 43:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 235 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: BRSTNOT04  
 (B) CLONE: 1969426

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43 :

Met	Ser	Pro	Thr	Leu	Ser	Ser	Ile	Thr	Gln	Gly	Val	Pro	Leu	Asp	5	10	15
Thr	Ser	Lys	Leu	Ser	Thr	Asp	Gln	Arg	Leu	Pro	Pro	Tyr	Pro	Tyr	20	25	30
Ser	Ser	Pro	Ser	Leu	Val	Leu	Pro	Thr	Gln	Pro	His	Thr	Pro	Lys	35	40	45
Ser	Leu	Gln	Gln	Pro	Gly	Leu	Pro	Ser	Gln	Ser	Cys	Ser	Val	Gln	50	55	60
Ser	Ser	Gly	Gly	Gln	Pro	Pro	Gly	Arg	Gln	Ser	His	Tyr	Gly	Thr	65	70	75
Pro	Tyr	Pro	Pro	Gly	Pro	Ser	Gly	His	Gly	Gln	Gln	Ser	Tyr	His	80	85	90
Arg	Pro	Met	Ser	Asp	Phe	Asn	Leu	Gly	Asn	Leu	Glu	Gln	Phe	Ser	95	100	105
Met	Glu	Ser	Pro	Ser	Ala	Ser	Leu	Val	Leu	Asp	Pro	Pro	Gly	Phe	110	115	120
Ser	Glu	Gly	Pro	Gly	Phe	Leu	Gly	Gly	Glu	Gly	Pro	Met	Gly	Gly	125	130	135
Pro	Gln	Asp	Pro	His	Thr	Phe	Asn	His	Gln	Asn	Leu	Thr	His	Cys	140	145	150
Ser	Arg	His	Gly	Ser	Gly	Pro	Asn	Ile	Ile	Leu	Thr	Gly	Asp	Ser	155	160	165
Ser	Pro	Gly	Phe	Ser	Lys	Glu	Ile	Ala	Ala	Ala	Leu	Ala	Gly	Val	170	175	180
Pro	Gly	Phe	Glu	Val	Ser	Ala	Ala	Gly	Leu	Glu	Leu	Gly	Leu	Gly	185	190	195
Leu	Glu	Asp	Glu	Leu	Arg	Met	Glu	Pro	Leu	Gly	Leu	Glu	Gly	Leu	200	205	210
Asn	Met	Leu	Ser	Asp	Pro	Cys	Ala	Leu	Leu	Pro	Asp	Pro	Ala	Val	215	220	225
Glu	Glu	Ser	Phe	Arg	Ser	Asp	Arg	Leu	Gln						230	235	

## (2) INFORMATION FOR SEQ ID NO: 44:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 203 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: UCMCL5T01

(B) CLONE: 1969948

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44 :

Met	Asn	Tyr	Phe	Pro	Leu	Ala	Pro	Phe	Asn	Gln	Leu	Leu	Gln	Lys	
				5					10					15	
Asp	Ile	Ile	Ser	Glu	Leu	Leu	Thr	Ser	Asp	Asp	Met	Lys	Asn	Ala	
				20					25					30	
Tyr	Lys	Leu	His	Thr	Leu	Asp	Thr	Cys	Leu	Lys	Leu	Asp	Asp	Thr	
				35					40					45	
Val	Tyr	Leu	Arg	Asp	Ile	Ala	Leu	Ser	Leu	Pro	Gln	Leu	Pro	Arg	
				50					55					60	
Glu	Leu	Pro	Ser	Ser	His	Thr	Asn	Ala	Lys	Val	Ala	Glu	Val	Leu	
				65					70					75	
Ser	Ser	Leu	Leu	Gly	Gly	Glu	Gly	His	Phe	Ser	Lys	Asp	Val	His	
				80					85					90	
Leu	Pro	His	Asn	Tyr	His	Ile	Asp	Phe	Glu	Ile	Arg	Met	Asp	Thr	
				95					100					105	
Asn	Arg	Asn	Gln	Val	Leu	Pro	Leu	Ser	Asp	Val	Asp	Thr	Thr	Ser	
				110					115					120	
Ala	Thr	Asp	Ile	Gln	Arg	Val	Ala	Val	Leu	Cys	Val	Ser	Arg	Ser	
				125					130					135	
Ala	Tyr	Cys	Leu	Gly	Ser	Ser	His	Pro	Arg	Gly	Phe	Leu	Ala	Met	
				140					145					150	
Lys	Met	Arg	His	Leu	Asn	Ala	Met	Gly	Phe	His	Val	Ile	Leu	Val	
				155					160					165	
Asn	Asn	Trp	Glu	Met	Asp	Lys	Leu	Glu	Met	Glu	Asp	Ala	Val	Thr	
				170					175					180	
Phe	Leu	Lys	Thr	Lys	Ile	Tyr	Ser	Val	Glu	Ala	Leu	Pro	Val	Ala	
				185					190					195	
Ala	Val	Asn	Val	Gln	Ser	Thr	Gln								
				200											

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 359 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGAST01
- (B) CLONE: 1988911

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45 :

Met	Glu	Arg	Gly	Asn	Val	Leu	Ser	Arg	Ala	Pro	Ser	Arg	Ala	His	
				5					10					15	
Gly	Thr	His	Phe	Gly	Asp	Asp	Arg	Phe	Glu	Asp	Leu	Glu	Glu	Ala	
				20					25					30	
Asn	Pro	Phe	Ser	Phe	Arg	Glu	Phe	Leu	Lys	Thr	Lys	Asn	Leu	Gly	
				35					40					45	
Leu	Ser	Lys	Glu	Asp	Pro	Ala	Ser	Arg	Ile	Tyr	Ala	Lys	Glu	Ala	
				50					55					60	
Ser	Arg	His	Ser	Leu	Gly	Leu	Asp	His	Asn	Ser	Pro	Pro	Ser	Gln	
				65					70					75	
Thr	Gly	Gly	Tyr	Gly	Leu	Glu	Tyr	Gln	Gln	Pro	Phe	Phe	Glu	Asp	

	80		85		90
Pro Thr Gly Ala Gly Asp Leu Leu Asp		Glu Glu Asp Glu Asp			
	95		100		105
Thr Gly Trp Ser Gly Ala Tyr Leu Pro		Ser Ala Ile Glu Gln Thr			
	110		115		120
His Pro Glu Arg Val Pro Ala Gly Thr		Ser Pro Cys Ser Thr Tyr			
	125		130		135
Leu Ser Phe Phe Ser Thr Pro Ser Glu		Leu Ala Gly Pro Glu Ser			
	140		145		150
Leu Pro Ser Trp Ala Leu Ser Asp Thr		Asp Ser Arg Val Ser Pro			
	155		160		165
Ala Ser Pro Ala Gly Ser Pro Ser Ala		Asp Phe Ala Val His Gly			
	170		175		180
Glu Ser Leu Gly Asp Arg His Leu Arg		Thr Leu Gln Ile Ser Tyr			
	185		190		195
Asp Ala Leu Lys Asp Glu Asn Ser Lys		Leu Arg Arg Lys Leu Asn			
	200		205		210
Glu Val Gln Ser Phe Ser Glu Ala Gln		Thr Glu Met Val Arg Thr			
	215		220		225
Leu Glu Arg Lys Leu Glu Ala Lys Met		Ile Lys Glu Glu Ser Asp			
	230		235		240
Tyr His Asp Leu Glu Ser Val Val Gln		Gln Val Glu Gln Asn Leu			
	245		250		255
Glu Leu Met Thr Lys Arg Ala Val Lys		Ala Glu Asn His Val Val			
	260		265		270
Lys Leu Lys Gln Glu Ile Ser Leu Leu		Gln Ala Gln Val Ser Asn			
	275		280		285
Phe Gln Arg Glu Asn Glu Ala Leu Arg		Cys Gly Gln Gly Ala Ser			
	290		295		300
Leu Thr Val Val Lys Gln Asn Ala Asp		Val Ala Leu Gln Asn Leu			
	305		310		315
Arg Val Val Met Asn Ser Ala Gln Ala		Ser Ile Lys Gln Leu Val			
	320		325		330
Ser Gly Ala Glu Thr Leu Asn Leu Val		Ala Glu Ile Leu Lys Ser			
	335		340		345
Ile Asp Arg Ile Ser Glu Val Lys Asp		Glu Glu Glu Asp Ser			
	350		355		

## (2) INFORMATION FOR SEQ ID NO: 46:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 150 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: OVARNOT03
- (B) CLONE: 2061561

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46 :

Met Gly Gly Lys Pro His Lys Glu Pro Arg Ala Lys Gly Pro Leu		
	5	10
Ser Ile Phe Tyr Pro Gly Ser Thr Ala Pro Val Ile Thr Gln Arg		
	20	25
Thr Pro Xaa Ala Ala Leu Lys Pro Pro Ile Lys Gly Ala Gly		
	35	40
		45

PF-0459 US

Pro	Thr	Ile	Ala	Pro	Ile	Lys	Gly	Xaa	Xaa	Asn	Phe	Gly	Lys	Arg	
				50					55					60	
Pro	Thr	Val	Thr	Xaa	Pro	Xaa	Trp	Xaa	Ile	Ser	Pro	Asn	Trp	Gly	
				65					70					75	
Lys	Arg	Gly	Xaa	Cys	Xaa	Xaa	Xaa	Gly	Ile	Lys	Trp	Val	Xaa	Pro	
				80					85					90	
Arg	Val	Ser	Gln	Ala	Arg	Thr	Phe	Lys	Thr	Thr	Ala	Asn	Glu	Leu	
				95					100					105	
Xaa	Phe	Xaa	Asp	Thr	Phe	Glu	Glu	Xaa	Xaa	Arg	Xaa	Xaa	His	Ala	
				110					115					120	
Xaa	Val	Ser	Xaa	Glu	Pro	Gln	Pro	Arg	Cys	Pro	Leu	Gly	Glu	Ser	
				125					130					135	
Arg	Ser	Leu	Gly	Ala	Ala	Val	Cys	Arg	Trp	Asp	Ser	Phe	Asp	Phe	
				140					145					150	

(2) INFORMATION FOR SEQ ID NO: 47:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 402 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: PANCNOT04  
(B) CLONE: 2084489

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47 :

Met	Pro	Pro	Val	Ser	Arg	Ser	Ser	Tyr	Ser	Glu	Asp	Ile	Val	Gly	
				5					10					15	
Ser	Arg	Arg	Arg	Arg	Arg	Ser	Ser	Ser	Gly	Ser	Pro	Pro	Ser	Pro	
				20					25					30	
Gln	Ser	Arg	Cys	Ser	Ser	Trp	Asp	Gly	Cys	Ser	Arg	Ser	His	Ser	
				35					40					45	
Arg	Gly	Arg	Glu	Gly	Leu	Arg	Pro	Pro	Trp	Ser	Glu	Leu	Asp	Val	
				50					55					60	
Gly	Ala	Leu	Tyr	Pro	Phe	Ser	Arg	Ser	Gly	Ser	Arg	Gly	Arg	Leu	
				65					70					75	
Pro	Arg	Phe	Arg	Asn	Tyr	Ala	Phe	Ala	Ser	Ser	Trp	Ser	Thr	Ser	
				80					85					90	
Tyr	Ser	Gly	Tyr	Arg	Tyr	His	Arg	His	Cys	Tyr	Ala	Glu	Glu	Arg	
				95					100					105	
Gln	Ser	Ala	Glu	Asp	Tyr	Glu	Lys	Glu	Glu	Ser	His	Arg	Gln	Arg	
				110					115					120	
Arg	Leu	Lys	Glu	Arg	Glu	Arg	Ile	Gly	Glu	Leu	Gly	Ala	Pro	Glu	
				125					130					135	
Val	Trp	Gly	Pro	Ser	Pro	Lys	Phe	Pro	Gln	Leu	Asp	Ser	Asp	Glu	
				140					145					150	
His	Thr	Pro	Val	Glu	Asp	Glu	Glu	Glu	Val	Thr	His	Gln	Lys	Ser	
				155					160					165	
Ser	Ser	Ser	Asp	Ser	Asn	Ser	Glu	Glu	His	Arg	Lys	Lys	Lys	Thr	
				170					175					180	
Ser	Arg	Ser	Arg	Asn	Lys	Lys	Lys	Arg	Lys	Asn	Lys	Ser	Ser	Lys	
				185					190					195	
Arg	Lys	His	Arg	Lys	Tyr	Ser	Asp	Ser	Asp	Ser	Asn	Ser	Glu	Ser	

Asp Thr Asn Ser	200	Asp Ser Asp Asp Asp	205	Lys Arg Val Lys	210
	215		220		225
Lys Lys Lys Lys	230	Lys Lys Lys Lys His	235	Lys Thr Lys Lys Lys	240
Asn Lys Lys Thr	245	Lys Lys Glu Ser Ser	250	Asp Ser Ser Cys Lys	255
Ser Glu Glu Asp	260	Leu Ser Glu Ala Thr	265	Trp Met Glu Gln Pro	270
Val Ala Asp Thr	275	Met Asp Leu Ile Gly	280	Pro Glu Ala Pro Ile	285
His Thr Ser Gln	290	Asp Glu Lys Pro Leu	295	Lys Tyr Gly His Ala	300
Leu Pro Gly Glu	305	Gly Ala Ala Met Ala	310	Glu Tyr Val Lys Ala	315
Lys Arg Ile Pro	320	Arg Arg Gly Glu Ile	325	Gly Leu Thr Ser Glu	330
Ile Gly Ser Phe	335	Glu Cys Ser Gly Tyr	340	Val Met Ser Gly Ser	345
His Arg Arg Met	350	Glu Ala Val Arg Leu	355	Arg Lys Glu Asn Gln	360
Tyr Ser Ala Asp	365	Glu Lys Arg Ala Leu	370	Ala Ser Phe Asn Gln	375
Glu Arg Arg Lys	380	Arg Glu Ser Lys Ile	385	Leu Ala Ser Phe Arg	390
Met Val His Lys	395	Lys Thr Lys Glu Lys	400	Asp Asp Lys	

## (2) INFORMATION FOR SEQ ID NO: 48:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 311 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: SPLNFET02
- (B) CLONE: 2203226

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48 :

Met His Pro Ala Gly	5	Leu Ala Ala Ala Ala	10	Ala Gly Thr Pro Arg	15
Leu Pro Ser Lys Arg	20	Arg Ile Pro Val Ser	25	Gln Pro Gly Met Ala	30
Asp Pro His Gln Leu	35	Phe Asp Asp Thr Ser	40	Ser Ala Gln Ser Arg	45
Gly Tyr Gly Ala Gln	50	Arg Ala Pro Gly Gly	55	Leu Ser Tyr Pro Ala	60
Ala Ser Pro Thr Pro	65	His Ala Ala Phe Leu	70	Ala Asp Pro Val Ser	75
Asn Met Ala Met Ala	80	Tyr Gly Ser Ser Leu	85	Ala Ala Gln Gly Lys	90
Glu Leu Val Asp Lys	95	Asn Ile Asp Arg Phe	100	Ile Pro Ile Thr Lys	105
Leu Lys Tyr Tyr Phe	110	Ala Val Asp Thr Met	115	Tyr Val Gly Arg Lys	120

PF-0459 US

Leu	Gly	Leu	Leu	Phe	Phe	Pro	Tyr	Leu	His	Gln	Asp	Trp	Glu	Val	
				125					130					135	
Gln	Tyr	Gln	Gln	Asp	Thr	Pro	Val	Ala	Pro	Arg	Phe	Asp	Val	Asn	
				140					145					150	
Ala	Pro	Asp	Leu	Tyr	Ile	Pro	Ala	Met	Ala	Phe	Ile	Thr	Tyr	Val	
				155					160					165	
Leu	Val	Ala	Gly	Leu	Ala	Leu	Gly	Thr	Gln	Asp	Arg	Phe	Ser	Pro	
				170					175					180	
Asp	Leu	Leu	Gly	Leu	Gln	Ala	Ser	Ser	Ala	Leu	Ala	Trp	Leu	Thr	
				185					190					195	
Leu	Glu	Val	Leu	Ala	Ile	Leu	Leu	Ser	Leu	Tyr	Leu	Val	Thr	Val	
				200					205					210	
Asn	Thr	Asp	Leu	Thr	Thr	Ile	Asp	Leu	Val	Ala	Phe	Leu	Gly	Tyr	
				215					220					225	
Lys	Tyr	Val	Gly	Met	Ile	Gly	Gly	Val	Leu	Met	Gly	Leu	Leu	Phe	
				230					235					240	
Gly	Lys	Ile	Gly	Tyr	Tyr	Leu	Val	Leu	Gly	Trp	Cys	Cys	Val	Ala	
				245					250					255	
Ile	Phe	Val	Phe	Met	Ile	Arg	Thr	Leu	Arg	Leu	Lys	Ile	Leu	Ala	
				260					265					270	
Asp	Ala	Ala	Ala	Glu	Gly	Val	Pro	Val	Arg	Gly	Ala	Arg	Asn	Gln	
				275					280					285	
Leu	Arg	Met	Tyr	Leu	Thr	Met	Ala	Val	Ala	Ala	Ala	Gln	Pro	Met	
				290					295					300	
Leu	Met	Tyr	Trp	Leu	Thr	Phe	His	Leu	Val	Arg					
				305					310						

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 316 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PROSNOT16
- (B) CLONE: 2232884

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49 :

Met	Ala	Ser	Ala	Asp	Glu	Leu	Thr	Phe	His	Glu	Phe	Glu	Glu	Ala	
				5					10					15	
Thr	Asn	Leu	Leu	Ala	Asp	Thr	Pro	Asp	Ala	Ala	Thr	Thr	Ser	Arg	
				20					25					30	
Ser	Asp	Gln	Leu	Thr	Pro	Gln	Gly	His	Val	Ala	Val	Ala	Val	Gly	
				35					40					45	
Ser	Gly	Gly	Ser	Tyr	Gly	Ala	Glu	Asp	Glu	Val	Glu	Glu	Glu	Ser	
				50					55					60	
Asp	Lys	Ala	Ala	Leu	Leu	Gln	Glu	Gln	Gln	Gln	Gln	Gln	Gln	Pro	
				65					70					75	
Gly	Phe	Trp	Thr	Phe	Ser	Tyr	Tyr	Gln	Ser	Phe	Phe	Asp	Val	Asp	
				80					85					90	
Thr	Ser	Gln	Val	Leu	Asp	Arg	Ile	Lys	Gly	Ser	Leu	Leu	Pro	Arg	
				95					100					105	
Pro	Gly	His	Asn	Phe	Val	Arg	His	His	Leu	Arg	Asn	Arg	Pro	Asp	
				110					115					120	
Leu	Tyr	Gly	Pro	Phe	Trp	Ile	Cys	Ala	Thr	Leu	Ala	Phe	Val	Leu	

	125		130		135
Ala Val Thr Gly	Asn Leu Thr Leu Val	Leu Ala Gln Arg Arg	Asp		
	140		145		150
Pro Ser Ile His	Tyr Ser Pro Gln Phe	His Lys Val Thr Val	Ala		
	155		160		165
Gly Ile Ser Ile	Tyr Cys Tyr Ala Trp	Leu Val Pro Leu Ala	Leu		
	170		175		180
Trp Gly Phe Leu	Arg Trp Arg Lys Gly	Val Gln Glu Arg Met	Gly		
	185		190		195
Pro Tyr Thr Phe	Leu Glu Thr Val Cys	Ile Tyr Gly Tyr Ser	Leu		
	200		205		210
Phe Val Phe Ile	Pro Met Val Val Leu	Trp Leu Ile Pro Val	Pro		
	215		220		225
Trp Leu Gln Trp	Phe Phe Gly Ala Leu	Ala Leu Gly Leu Ser	Ala		
	230		235		240
Ala Gly Leu Val	Phe Thr Leu Trp Pro	Val Val Arg Glu Asp	Thr		
	245		250		255
Arg Leu Val Ala	Thr Val Leu Leu Ser	Val Val Val Leu Leu	His		
	260		265		270
Ala Leu Leu Ala	Met Gly Cys Lys Leu	Tyr Phe Phe Gln Ser	Leu		
	275		280		285
Pro Pro Glu Asn	Val Ala Pro Pro Pro	Gln Ile Thr Ser Leu	Pro		
	290		295		300
Ser Asn Ile Ala	Leu Ser Pro Thr Leu	Pro Gln Ser Leu Ala	Pro		
	305		310		315
Ser					

## (2) INFORMATION FOR SEQ ID NO: 50:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 346 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: COLNNOT11
- (B) CLONE: 2328134

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50 :

Met Thr Pro Arg Thr	Trp Trp Pro Arg	Pro Ala Gly Trp Gly Thr	
	5	10	15
Cys Arg Ala Ala Gly	Trp Pro Arg Ser Val	Pro Trp Ala Arg Thr	
	20	25	30
Ala Ala Ser Leu Val	Phe Val Pro Thr Arg	Arg Arg Ser Gly Pro	
	35	40	45
Ser Gly Thr Ala Ser	Val Ala Ala Met Ala	Tyr His Ser Gly Tyr	
	50	55	60
Gly Ala His Gly Ser	Lys His Arg Ala Arg	Ala Ala Pro Asp Pro	
	65	70	75
Pro Pro Leu Phe Asp	Asp Thr Ser Gly Gly	Tyr Ser Ser Gln Pro	
	80	85	90
Gly Gly Tyr Pro Ala	Thr Gly Ala Asp Val	Ala Phe Ser Val Asn	
	95	100	105
His Leu Leu Gly Asp	Pro Met Ala Asn Val	Ala Met Ala Tyr Gly	
	110	115	120
Ser Ser Ile Ala Ser	His Gly Lys Asp Met	Val His Lys Glu Leu	
	125	130	135



PF-0459 US

His Arg Phe Val	Ser Val Ser Lys Leu	Lys Tyr Phe Phe Ala Val	140	145	150
Asp Thr Ala Tyr	Val Ala Lys Lys Leu	Gly Leu Leu Val Phe Pro	155	160	165
Tyr Thr His Gln	Asn Trp Glu Val Gln	Tyr Ser Arg Asp Ala Pro	170	175	180
Leu Pro Pro Arg	Gln Asp Leu Asn Ala	Pro Asp Leu Tyr Ile Pro	185	190	195
Thr Met Ala Phe	Ile Thr Tyr Val Leu	Leu Ala Gly Met Ala Leu	200	205	210
Gly Ile Gln Lys	Arg Phe Ser Pro Glu	Val Leu Gly Leu Cys Ala	215	220	225
Ser Thr Ala Leu	Val Trp Val Val Met	Glu Val Leu Ala Leu Leu	230	235	240
Leu Gly Leu Tyr	Leu Ala Thr Val Arg	Ser Asp Leu Ser Thr Phe	245	250	255
His Leu Leu Ala	Tyr Ser Gly Tyr Lys	Tyr Val Gly Met Ile Leu	260	265	270
Ser Val Leu Thr	Gly Leu Leu Phe Gly	Ser Asp Gly Tyr Tyr Val	275	280	285
Ala Leu Ala Trp	Thr Ser Ser Ala Leu	Met Tyr Phe Ile Val Arg	290	295	300
Ser Leu Arg Thr	Ala Ala Leu Gly Pro	Asp Ser Met Gly Gly Pro	305	310	315
Val Pro Arg Gln	Arg Leu Gln Leu Tyr	Leu Thr Leu Gly Ala Ala	320	325	330
Ala Phe Gln Pro	Leu Ile Ile Tyr Trp	Leu Thr Phe His Leu Val	335	340	345
Arg					

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 299 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: ISLTNOT01
- (B) CLONE: 2382718

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51 :

Met Gly Thr Lys Ala Gln Val Glu Arg Lys Leu Leu Cys Leu Phe	5	10	15
Ile Leu Ala Ile Leu Leu Cys Ser Leu Ala Leu Gly Ser Val Thr	20	25	30
Val His Ser Ser Glu Pro Glu Val Arg Ile Pro Glu Asn Asn Pro	35	40	45
Val Lys Leu Ser Cys Ala Tyr Ser Gly Phe Ser Ser Pro Arg Val	50	55	60
Glu Trp Lys Phe Asp Gln Gly Asp Thr Thr Arg Leu Val Cys Tyr	65	70	75
Asn Asn Lys Ile Thr Ala Ser Tyr Glu Asp Arg Val Thr Phe Leu	80	85	90
Pro Thr Gly Ile Thr Phe Lys Ser Val Thr Arg Glu Asp Thr Gly	95	100	105
Thr Tyr Thr Cys Met Val Ser Glu Glu Gly Gly Asn Ser Tyr Gly			

	110		115		120
Glu Val Lys Val Lys	Leu Ile Val Leu	Val Pro Pro Ser Lys	Pro		
	125		130		135
Thr Val Asn Ile Pro	Ser Ser Ala Thr	Ile Gly Asn Arg Ala	Val		
	140		145		150
Leu Thr Cys Ser Glu	Gln Asp Gly Ser	Pro Pro Ser Glu Tyr	Thr		
	155		160		165
Trp Phe Lys Asp Gly	Ile Val Met Pro	Thr Asn Pro Lys Ser	Thr		
	170		175		180
Arg Ala Phe Ser Asn	Ser Ser Tyr Val	Leu Asn Pro Thr Thr	Gly		
	185		190		195
Glu Leu Val Phe Asp	Pro Leu Ser Ala	Ser Asp Thr Gly Glu	Tyr		
	200		205		210
Ser Cys Glu Ala Arg	Asn Gly Tyr Gly	Thr Pro Met Thr Ser	Asn		
	215		220		225
Ala Val Arg Met Glu	Ala Val Glu Arg	Asn Val Gly Val Ile	Val		
	230		235		240
Ala Ala Val Leu Val	Thr Leu Ile Leu	Leu Gly Ile Leu Val	Phe		
	245		250		255
Gly Ile Trp Phe Ala	Tyr Ser Arg Gly	His Phe Asp Arg Thr	Lys		
	260		265		270
Lys Gly Thr Ser Ser	Lys Lys Val Ile	Tyr Ser Gln Pro Ser	Ala		
	275		280		285
Arg Ser Glu Gly Glu	Phe Lys Gln Thr	Ser Ser Phe Leu Val			
	290		295		

## (2) INFORMATION FOR SEQ ID NO: 52:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 351 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: ENDANOT01
- (B) CLONE: 2452208

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52 :

Met Ala Ser Thr Gly	Ser Gln Ala Ser	Asp Ile Asp Glu Ile	Phe
	5	10	15
Gly Phe Phe Asn Asp	Gly Glu Pro Pro	Thr Lys Lys Pro Arg	Lys
	20	25	30
Leu Leu Pro Ser Leu	Lys Thr Lys Lys	Pro Arg Glu Leu Val	Leu
	35	40	45
Val Ile Gly Thr Gly	Ile Ser Ala Ala	Val Ala Pro Gln Val	Pro
	50	55	60
Ala Leu Lys Ser Trp	Lys Gly Leu Ile	Gln Ala Leu Leu Asp	Ala
	65	70	75
Ala Ile Asp Phe Asp	Leu Leu Glu Asp	Glu Glu Ser Lys Lys	Phe
	80	85	90
Gln Lys Cys Leu His	Glu Asp Lys Asn	Leu Val His Val Ala	His
	95	100	105
Asp Leu Ile Gln Lys	Leu Ser Pro Arg	Thr Ser Asn Val Arg	Ser
	110	115	120
Thr Phe Phe Lys Asp	Cys Leu Tyr Glu	Val Phe Asp Asp Leu	Glu
	125	130	135

Ser	Lys	Met	Glu	Asp	Ser	Gly	Lys	Gln	Leu	Leu	Gln	Ser	Val	Leu	
				140					145					150	
His	Leu	Met	Glu	Asn	Gly	Ala	Leu	Val	Leu	Thr	Thr	Asn	Phe	Asp	
				155					160					165	
Asn	Leu	Leu	Glu	Leu	Tyr	Ala	Ala	Asp	Gln	Gly	Lys	Gln	Leu	Glu	
				170					175					180	
Ser	Leu	Asp	Leu	Thr	Asp	Glu	Lys	Lys	Val	Leu	Glu	Trp	Ala	Gln	
				185					190					195	
Glu	Lys	Arg	Lys	Leu	Ser	Val	Leu	His	Ile	His	Gly	Val	Tyr	Thr	
				200					205					210	
Asn	Pro	Ser	Gly	Ile	Val	Leu	His	Pro	Ala	Gly	Tyr	Gln	Asn	Val	
				215					220					225	
Leu	Arg	Asn	Thr	Glu	Val	Met	Arg	Glu	Ile	Gln	Lys	Leu	Tyr	Glu	
				230					235					240	
Asn	Lys	Ser	Phe	Leu	Phe	Leu	Gly	Cys	Gly	Trp	Thr	Val	Asp	Asp	
				245					250					255	
Thr	Thr	Phe	Gln	Ala	Leu	Phe	Leu	Glu	Ala	Val	Lys	His	Lys	Ser	
				260					265					270	
Asp	Leu	Glu	His	Phe	Met	Leu	Val	Arg	Arg	Gly	Asp	Val	Asp	Glu	
				275					280					285	
Phe	Lys	Lys	Leu	Arg	Glu	Asn	Met	Leu	Asp	Lys	Gly	Ile	Lys	Val	
				290					295					300	
Ile	Ser	Tyr	Gly	Asp	Asp	Tyr	Ala	Asp	Leu	Pro	Glu	Tyr	Phe	Lys	
				305					310					315	
Arg	Leu	Thr	Cys	Glu	Ile	Ser	Thr	Arg	Gly	Thr	Ser	Ala	Gly	Met	
				320					325					330	
Val	Arg	Glu	Gly	Gln	Leu	Asn	Gly	Ser	Ser	Ala	Ala	His	Ser	Glu	
				335					340					345	
Ile	Arg	Gly	Cys	Ser	Thr										
				350											

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 662 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: ENDANOT01
- (B) CLONE: 2457825

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53 :

Met	Thr	Ala	Lys	Lys	Gln	Cys	Leu	Leu	Arg	Leu	Gly	Val	Leu	Arg	
				5					10					15	
Gln	Asp	Trp	Pro	Asp	Thr	Asn	Arg	Leu	Leu	Gly	Ser	Ala	Asn	Val	
				20					25					30	
Val	Pro	Glu	Ala	Leu	Gln	Arg	Phe	Thr	Arg	Ala	Ala	Ala	Asp	Phe	
				35					40					45	
Ala	Thr	His	Gly	Lys	Leu	Gly	Lys	Leu	Glu	Phe	Ala	Gln	Asp	Ala	
				50					55					60	
His	Gly	Gln	Pro	Asp	Val	Ser	Ala	Phe	Asp	Phe	Thr	Ser	Met	Met	
				65					70					75	
Arg	Ala	Glu	Ser	Ser	Ala	Arg	Val	Gln	Glu	Lys	His	Gly	Ala	Arg	
				80					85					90	
Leu	Leu	Leu	Gly	Leu	Val	Gly	Asp	Cys	Leu	Val	Glu	Pro	Phe	Trp	

Pro Leu Gly Thr	95	Gly Val Ala Arg Gly	100	Phe Leu Ala Ala Phe	105
Ala Ala Trp Met	110	Val Lys Arg Trp Ala	115	Glu Gly Ala Glu Ser	120
Glu Val Leu Ala	125	Glu Arg Glu Ser Leu	130	Tyr Gln Leu Leu Ser	135
Thr Ser Pro Glu	140	Asn Met His Arg Asn	145	Val Ala Gln Tyr Gly	150
Asp Pro Ala Thr	155	Arg Tyr Pro Asn Leu	160	Asn Leu Arg Ala Val	165
Pro Asn Gln Val	170	Arg Asp Leu Tyr Asp	175	Val Leu Ala Lys Glu	180
Val Gln Arg Asp	185	Asn Asp Lys Thr Asp	190	Thr Gly Met Pro Ala	195
Gly Ser Ala Gly	200	Thr Gln Glu Glu Leu	205	Leu Arg Trp Cys Gln	210
Gln Thr Ala Gly	215	Tyr Pro Gly Val His	220	Val Ser Asp Leu Ser	225
Ser Trp Ala Asp	230	Gly Leu Ala Leu Cys	235	Ala Leu Val Tyr Arg	240
Gln Pro Gly Leu	245	Leu Glu Pro Ser Glu	250	Leu Gln Gly Leu Gly	255
Leu Glu Ala Thr	260	Ala Trp Ala Leu Lys	265	Val Ala Glu Asn Glu	270
Gly Ile Thr Pro	275	Val Val Ser Ala Gln	280	Ala Val Val Ala Gly	285
Asp Pro Leu Gly	290	Leu Ile Ala Tyr Leu	295	Ser His Phe His Ser	300
Phe Lys Ser Met	305	Ala His Ser Pro Gly	310	Pro Val Ser Gln Ala	315
Pro Gly Thr Ser	320	Ser Ala Val Leu Phe	325	Leu Ser Lys Leu Gln	330
Thr Leu Gln Arg	335	Ser Arg Ala Lys Glu	340	Asn Ala Glu Asp Ala	345
Gly Lys Lys Leu	350	Arg Leu Glu Met Glu	355	Ala Glu Thr Pro Ser	360
Glu Val Pro Pro	365	Asp Pro Glu Pro Gly	370	Val Pro Leu Thr Pro	375
Ser Gln His Gln	380	Glu Ala Gly Ala Gly	385	Asp Leu Cys Ala Leu	390
Gly Glu His Leu	395	Tyr Val Leu Glu Arg	400	Leu Cys Val Asn Gly	405
Phe Phe His Arg	410	Ser Cys Phe Arg Cys	415	His Thr Cys Glu Ala	420
Leu Trp Pro Gly	425	Gly Tyr Glu Gln His	430	Pro Gly Ser Arg Thr	435
Gln Phe Phe Phe	440	Ser Ala Leu Val Ala	445	Met Glu Lys Glu Glu	450
Glu Ser Pro Phe	455	Ser Ser Glu Glu Glu	460	Glu Glu Asp Val Pro	465
Asp Ser Asp Val	470	Glu Gln Ala Leu Gln	475	Thr Phe Ala Lys Thr	480
Gly Thr Met Asn	485	Asn Tyr Pro Thr Trp	490	Arg Arg Thr Leu Leu	495
Arg Ala Lys Glu	500	Glu Glu Met Lys Arg	505	Phe Cys Lys Ala Gln	510
Ile Gln Arg Arg	515	Leu Asn Glu Ile Glu	520	Ala Ala Leu Arg Glu	525
Glu Ala Glu Gly	530	Val Lys Leu Glu Leu	535	Ala Leu Arg Arg Gln	540
	545		550		555

PF-0459 US

Ser	Ser	Pro	Glu	Gln	Gln	Lys	Lys	Leu	Trp	Val	Gly	Gln	Leu	Leu
				560					565					570
Gln	Leu	Val	Asp	Lys	Lys	Asn	Ser	Leu	Val	Ala	Glu	Glu	Ala	Glu
				575					580					585
Leu	Met	Ile	Thr	Val	Gln	Glu	Leu	Asn	Leu	Glu	Glu	Lys	Gln	Trp
				590					595					600
Gln	Leu	Asp	Gln	Glu	Leu	Arg	Gly	Tyr	Met	Asn	Arg	Glu	Glu	Asn
				605					610					615
Leu	Lys	Thr	Ala	Ala	Asp	Arg	Gln	Ala	Glu	Asp	Gln	Val	Leu	Arg
				620					625					630
Lys	Leu	Val	Asp	Leu	Val	Asn	Gln	Arg	Asp	Ala	Leu	Ile	Arg	Phe
				635					640					645
Gln	Glu	Glu	Arg	Arg	Leu	Ser	Glu	Leu	Ala	Leu	Gly	Thr	Gly	Ala
				650					655					660
Gln	Gly													

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 115 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: THP1NOT03
- (B) CLONE: 2470740

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54 :

Met	Ala	Ser	Trp	Pro	Ala	Ser	Pro	Leu	Gln	Trp	Gly	Pro	Pro	Leu
				5					10					15
Ala	Ser	Cys	Pro	Ser	Cys	Cys	Cys	Cys	Cys	Phe	His	Cys	Trp	Gln
				20					25					30
Pro	Arg	Val	Gly	Val	Ala	Cys	Arg	Gln	Arg	Cys	Trp	Pro	Leu	Arg
				35					40					45
Trp	Gly	Trp	Trp	Val	Trp	Gly	Pro	Pro	Thr	Cys	Ser	Phe	Val	Gln
				50					55					60
Pro	Cys	Thr	Cys	Pro	Pro	Val	Phe	Ser	Tyr	Ser	Trp	Pro	Arg	Val
				65					70					75
Pro	His	Trp	Gly	Pro	Ser	Trp	Xaa	Met	Ser	Trp	Arg	Arg	Arg	Leu
				80					85					90
Met	Gly	Val	Pro	Leu	Gly	Leu	Trp	Asn	Cys	Leu	Val	Leu	Lys	Leu
				95					100					105
Xaa	Gln	Gly	Leu	Ala	Pro	Thr	Ser	Gly	Gly					
				110					115					

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 157 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: SMCANOT01  
(B) CLONE: 2479092

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55 :

Met	Glu	Ala	Leu	Arg	Arg	Ala	His	Glu	Val	Ala	Leu	Arg	Leu	Leu	
				5					10						15
Leu	Cys	Arg	Pro	Trp	Ala	Ser	Arg	Ala	Ala	Ala	Arg	Pro	Lys	Pro	
				20					25						30
Ser	Ala	Ser	Glu	Val	Leu	Thr	Arg	His	Leu	Leu	Gln	Arg	Arg	Leu	
				35					40						45
Pro	His	Trp	Thr	Ser	Phe	Cys	Val	Pro	Tyr	Ser	Ala	Val	Arg	Asn	
				50					55						60
Asp	Gln	Phe	Gly	Leu	Ser	His	Phe	Asn	Trp	Pro	Val	Gln	Gly	Ala	
				65					70						75
Asn	Tyr	His	Val	Leu	Arg	Thr	Gly	Cys	Phe	Pro	Phe	Ile	Lys	Tyr	
				80					85						90
His	Cys	Ser	Lys	Ala	Pro	Trp	Gln	Asp	Leu	Ala	Arg	Gln	Asn	Arg	
				95					100						105
Phe	Phe	Thr	Ala	Leu	Lys	Val	Val	Asn	Leu	Gly	Ile	Pro	Thr	Leu	
				110					115						120
Leu	Tyr	Gly	Leu	Gly	Ser	Trp	Leu	Phe	Ala	Arg	Val	Thr	Glu	Thr	
				125					130						135
Val	His	Thr	Ser	Tyr	Gly	Pro	Ile	Thr	Val	Tyr	Phe	Leu	Asn	Lys	
				140					145						150
Glu	Asp	Glu	Gly	Ala	Met	Tyr									
				155											

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 197 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: SMCANOT01
- (B) CLONE: 2480544

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56 :

Met	Pro	Pro	Ala	Gly	Leu	Arg	Arg	Ala	Ala	Pro	Leu	Thr	Ala	Ile	
				5					10						15
Ala	Leu	Leu	Val	Leu	Gly	Ala	Pro	Leu	Val	Leu	Ala	Gly	Glu	Asp	
				20					25						30
Cys	Leu	Trp	Tyr	Leu	Asp	Arg	Asn	Gly	Ser	Trp	His	Pro	Gly	Phe	
				35					40						45
Asn	Cys	Glu	Phe	Phe	Thr	Phe	Cys	Cys	Gly	Thr	Cys	Tyr	His	Arg	
				50					55						60
Tyr	Cys	Cys	Arg	Asp	Leu	Thr	Leu	Leu	Ile	Thr	Glu	Arg	Gln	Gln	
				65					70						75
Lys	His	Cys	Leu	Ala	Phe	Ser	Pro	Lys	Thr	Ile	Ala	Gly	Ile	Ala	
				80					85						90
Ser	Ala	Val	Ile	Leu	Phe	Val	Ala	Val	Val	Ala	Thr	Thr	Ile	Cys	
				95					100						105
Cys	Phe	Leu	Cys	Ser	Cys	Cys	Tyr	Leu	Tyr	Arg	Arg	Arg	Gln	Gln	
				110					115						120

PF-0459 US

Leu	Gln	Ser	Pro	Phe	Glu	Gly	Gln	Glu	Ile	Pro	Met	Thr	Gly	Ile
				125					130					135
Pro	Val	Gln	Pro	Val	Tyr	Pro	Tyr	Pro	Gln	Asp	Pro	Lys	Ala	Gly
				140					145					150
Pro	Ala	Pro	Pro	Gln	Pro	Gly	Phe	Met	Tyr	Pro	Pro	Ser	Gly	Pro
				155					160					165
Ala	Pro	Gln	Tyr	Pro	Leu	Tyr	Pro	Ala	Gly	Pro	Pro	Val	Tyr	Asn
				170					175					180
Pro	Ala	Ala	Pro	Pro	Pro	Tyr	Met	Pro	Pro	Gln	Pro	Ser	Tyr	Pro
				185					190					195
Gly	Ala													

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 245 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAITUT21
- (B) CLONE: 2518547

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57 :

Met	Gly	Gly	Ala	Ser	Arg	Arg	Val	Glu	Ser	Gly	Ala	Trp	Ala	Tyr
				5					10					15
Leu	Ser	Pro	Leu	Val	Leu	Arg	Lys	Glu	Leu	Glu	Ser	Leu	Val	Glu
				20					25					30
Asn	Glu	Gly	Ser	Glu	Val	Leu	Ala	Leu	Pro	Glu	Leu	Pro	Ser	Ala
				35					40					45
His	Pro	Ile	Ile	Phe	Trp	Asn	Leu	Leu	Trp	Tyr	Phe	Gln	Arg	Leu
				50					55					60
Arg	Leu	Pro	Ser	Ile	Leu	Pro	Gly	Leu	Val	Leu	Ala	Ser	Cys	Asp
				65					70					75
Gly	Pro	Ser	His	Ser	Gln	Ala	Pro	Ser	Pro	Trp	Leu	Thr	Pro	Asp
				80					85					90
Pro	Ala	Ser	Val	Gln	Val	Arg	Leu	Leu	Trp	Asp	Val	Leu	Thr	Pro
				95					100					105
Asp	Pro	Asn	Ser	Cys	Pro	Pro	Leu	Tyr	Val	Leu	Trp	Arg	Val	His
				110					115					120
Ser	Gln	Ile	Pro	Gln	Arg	Val	Val	Trp	Pro	Gly	Pro	Val	Pro	Ala
				125					130					135
Ser	Leu	Ser	Leu	Ala	Leu	Leu	Glu	Ser	Val	Leu	Arg	His	Val	Gly
				140					145					150
Leu	Asn	Glu	Val	His	Lys	Ala	Val	Gly	Leu	Leu	Leu	Glu	Thr	Leu
				155					160					165
Gly	Pro	Pro	Pro	Thr	Gly	Leu	His	Leu	Gln	Arg	Gly	Ile	Tyr	Arg
				170					175					180
Glu	Ile	Leu	Phe	Leu	Thr	Met	Ala	Ala	Leu	Gly	Lys	Asp	His	Val
				185					190					195
Asp	Ile	Val	Ala	Phe	Asp	Lys	Lys	Tyr	Lys	Ser	Ala	Phe	Asn	Lys
				200					205					210
Leu	Ala	Ser	Ser	Met	Gly	Lys	Glu	Glu	Leu	Arg	His	Arg	Arg	Ala
				215					220					225
Gln	Met	Pro	Thr	Pro	Lys	Ala	Ile	Asp	Cys	Arg	Lys	Cys	Phe	Gly
				230					235					240

PF-0459 US

Ala Pro Pro Glu Cys  
245

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 310 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: GBLANOT02  
(B) CLONE: 2530650

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58 :

Met Leu Leu Pro Gln Leu Cys Trp Leu Pro Leu Leu Ala Gly Leu  
5 10 15  
Leu Pro Pro Val Pro Ala Gln Lys Phe Ser Ala Leu Thr Phe Leu  
20 25 30  
Arg Val Asp Gln Asp Lys Asp Lys Asp Cys Ser Leu Asp Cys Ala  
35 40 45  
Gly Ser Pro Gln Lys Pro Leu Cys Ala Ser Asp Gly Arg Thr Phe  
50 55 60  
Leu Ser Arg Cys Glu Phe Gln Arg Ala Lys Cys Lys Asp Pro Gln  
65 70 75  
Leu Glu Ile Ala Tyr Arg Gly Asn Cys Lys Asp Val Ser Arg Cys  
80 85 90  
Val Ala Glu Arg Lys Tyr Thr Gln Glu Gln Ala Arg Lys Glu Phe  
95 100 105  
Gln Gln Val Phe Ile Pro Glu Cys Asn Asp Asp Gly Thr Tyr Ser  
110 115 120  
Gln Val Gln Cys His Ser Tyr Thr Gly Tyr Cys Trp Cys Val Thr  
125 130 135  
Pro Asn Gly Arg Pro Ile Ser Gly Thr Ala Val Ala His Lys Thr  
140 145 150  
Pro Arg Cys Pro Gly Ser Val Asn Glu Lys Leu Pro Gln Arg Glu  
155 160 165  
Gly Thr Gly Lys Thr Asp Asp Ala Ala Ala Pro Ala Leu Glu Thr  
170 175 180  
Gln Pro Gln Gly Asp Glu Glu Asp Ile Ala Ser Arg Tyr Pro Thr  
185 190 195  
Leu Trp Thr Glu Gln Val Lys Ser Arg Gln Asn Lys Thr Asn Lys  
200 205 210  
Asn Ser Val Ser Ser Cys Asp Gln Glu His Gln Ser Ala Leu Glu  
215 220 225  
Glu Ala Lys Gln Pro Lys Asn Asp Asn Val Val Ile Pro Glu Cys  
230 235 240  
Ala His Gly Gly Leu Tyr Lys Pro Val Gln Cys His Pro Ser Thr  
245 250 255  
Gly Tyr Cys Trp Cys Val Leu Val Asp Thr Gly Arg Pro Ile Pro  
260 265 270  
Gly Thr Ser Thr Arg Tyr Glu Gln Pro Lys Cys Asp Asn Thr Gly  
275 280 285  
Gln Gly Pro Pro Ser Gln Ser Pro Gly Pro Val Gln Gly Pro Pro  
290 295 300  
Ala Thr Arg Leu Ser Gly Cys Gln Lys Ala



(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 256 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59 :

(i) SEQUENCE CHARACTERISTICS:

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: LUNGTUT11  
(B) CLONE: 2746976

Met	Gln	Phe	Met	Leu	Leu	Phe	Ser	Arg	Gln	Gly	Lys	Leu	Arg	Leu
				5					10					15
Gln	Lys	Trp	Tyr	Val	Pro	Leu	Ser	Asp	Lys	Glu	Lys	Arg	Lys	Ile
				20					25					30
Thr	Arg	Glu	Leu	Val	Gln	Thr	Val	Leu	Ala	Arg	Lys	Pro	Lys	Met
				35					40					45
Cys	Ser	Phe	Leu	Glu	Trp	Arg	Asp	Leu	Lys	Ile	Val	Tyr	Lys	Arg
				50					55					60
Tyr	Ala	Ser	Leu	Tyr	Phe	Cys	Cys	Ala	Ile	Glu	Asp	Gln	Asp	Asn
				65					70					75
Glu	Leu	Ile	Thr	Leu	Glu	Ile	Ile	His	Arg	Tyr	Val	Glu	Leu	Leu
				80					85					90
Asp	Lys	Tyr	Phe	Gly	Ser	Val	Cys	Glu	Leu	Asp	Ile	Ile	Phe	Asn
				95					100					105
Phe	Glu	Lys	Ala	Tyr	Phe	Ile	Leu	Asp	Glu	Phe	Leu	Leu	Gly	Gly
				110					115					120
Glu	Val	Gln	Glu	Thr	Ser	Lys	Lys	Asn	Val	Leu	Lys	Ala	Ile	Glu
				125					130					135
Gln	Ala	Asp	Leu	Leu	Gln	Glu	Asp	Ala	Lys	Glu	Ala	Glu	Thr	Pro
				140					145					150
Arg	Ser	Val	Leu	Glu	Glu	Ile	Gly	Leu	Thr					
				155					160					

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 341 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61 :

173

				80					85					90
Trp	Trp	Lys	Asn	Ala	Gln	Gly	Lys	Asp	Gln	Met	Trp	Tyr	Glu	Asp
				95					100					105
Ala	Leu	Ala	Ser	Ser	His	Pro	Ile	Ile	Leu	Tyr	Leu	His	Gly	Asn
				110					115					120
Ala	Gly	Thr	Arg	Gly	Gly	Asp	His	Arg	Val	Glu	Leu	Tyr	Lys	Val
				125					130					135
Leu	Ser	Ser	Leu	Gly	Tyr	His	Val	Val	Thr	Phe	Asp	Tyr	Arg	Gly
				140					145					150
Trp	Gly	Asp	Ser	Val	Gly	Thr	Pro	Ser	Glu	Arg	Gly	Met	Thr	Tyr
				155					160					165
Asp	Ala	Leu	His	Val	Phe	Asp	Trp	Ile	Lys	Ala	Arg	Ser	Gly	Asp
				170					175					180
Asn	Pro	Val	Tyr	Ile	Trp	Gly	His	Ser	Leu	Gly	Thr	Gly	Val	Ala
				185					190					195
Thr	Asn	Leu	Val	Arg	Arg	Leu	Cys	Glu	Arg	Glu	Thr	Pro	Pro	Asp
				200					205					210
Ala	Leu	Ile	Leu	Glu	Ser	Pro	Phe	Thr	Asn	Ile	Arg	Glu	Glu	Ala
				215					220					225
Lys	Ser	His	Pro	Phe	Ser	Val	Ile	Tyr	Arg	Tyr	Phe	Pro	Gly	Phe
				230					235					240
Asp	Trp	Phe	Phe	Leu	Asp	Pro	Ile	Thr	Ser	Ser	Gly	Ile	Lys	Phe
				245					250					255
Ala	Asn	Asp	Glu	Asn	Val	Lys	His	Ile	Ser	Cys	Pro	Leu	Leu	Ile
				260					265					270
Leu	His	Ala	Glu	Asp	Asp	Pro	Val	Val	Pro	Phe	Gln	Leu	Gly	Arg
				275					280					285
Lys	Leu	Tyr	Ser	Ile	Ala	Ala	Pro	Ala	Arg	Ser	Phe	Arg	Asp	Phe
				290					295					300
Lys	Val	Gln	Phe	Val	Pro	Phe	His	Ser	Asp	Leu	Gly	Tyr	Arg	His
				305					310					315
Lys	Tyr	Ile	Tyr	Lys	Ser	Pro	Glu	Leu	Pro	Arg	Ile	Leu	Arg	Glu
				320					325					330
Phe	Leu	Gly	Lys	Ser	Glu	Pro	Glu	His	Gln	His				
				335					340					

## (2) INFORMATION FOR SEQ ID NO: 62:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 430 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: OVARTUT03
- (B) CLONE: 2781553

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62 :

Met	Ala	Glu	Gly	Glu	Asp	Val	Gly	Trp	Trp	Arg	Ser	Trp	Leu	Gln
				5					10					15
Gln	Ser	Tyr	Gln	Ala	Val	Lys	Glu	Lys	Ser	Ser	Glu	Ala	Leu	Glu
				20					25					30
Phe	Met	Lys	Arg	Asp	Leu	Thr	Glu	Phe	Thr	Gln	Val	Val	Gln	His
				35					40					45
Asp	Thr	Ala	Cys	Thr	Ile	Ala	Ala	Thr	Ala	Ser	Val	Val	Lys	Glu
				50					55					60

PF-0459 US

Lys	Leu	Ala	Thr	Glu	Gly	Ser	Ser	Gly	Ala	Thr	Glu	Lys	Met	Lys	65	70	75
Lys	Gly	Leu	Ser	Asp	Phe	Leu	Gly	Val	Ile	Ser	Asp	Thr	Phe	Ala	80	85	90
Pro	Ser	Pro	Asp	Lys	Thr	Ile	Asp	Cys	Asp	Val	Ile	Thr	Leu	Met	95	100	105
Gly	Thr	Pro	Ser	Gly	Thr	Ala	Glu	Pro	Tyr	Asp	Gly	Thr	Lys	Ala	110	115	120
Arg	Leu	Tyr	Ser	Leu	Gln	Ser	Asp	Pro	Ala	Thr	Tyr	Cys	Asn	Glu	125	130	135
Pro	Asp	Gly	Pro	Pro	Glu	Leu	Phe	Asp	Ala	Trp	Leu	Ser	Gln	Phe	140	145	150
Cys	Leu	Glu	Glu	Lys	Lys	Gly	Glu	Ile	Ser	Glu	Leu	Leu	Val	Gly	155	160	165
Ser	Pro	Ser	Ile	Arg	Ala	Leu	Tyr	Thr	Lys	Met	Val	Pro	Ala	Ala	170	175	180
Val	Ser	His	Ser	Glu	Phe	Trp	His	Arg	Tyr	Phe	Tyr	Lys	Val	His	185	190	195
Gln	Leu	Glu	Gln	Glu	Gln	Ala	Arg	Arg	Asp	Ala	Leu	Lys	Gln	Arg	200	205	210
Ala	Glu	Gln	Ser	Ile	Ser	Glu	Glu	Pro	Gly	Trp	Glu	Glu	Glu	Glu	215	220	225
Glu	Glu	Leu	Met	Gly	Ile	Ser	Pro	Ile	Ser	Pro	Lys	Glu	Ala	Lys	230	235	240
Val	Pro	Val	Ala	Lys	Ile	Ser	Thr	Phe	Pro	Glu	Gly	Glu	Pro	Gly	245	250	255
Pro	Gln	Ser	Pro	Cys	Glu	Glu	Asn	Leu	Val	Thr	Ser	Val	Glu	Pro	260	265	270
Pro	Ala	Glu	Val	Thr	Pro	Ser	Glu	Ser	Ser	Glu	Ser	Ile	Ser	Leu	275	280	285
Val	Thr	Gln	Ile	Ala	Asn	Pro	Ala	Thr	Ala	Pro	Glu	Ala	Arg	Val	290	295	300
Leu	Pro	Lys	Asp	Leu	Ser	Gln	Lys	Leu	Leu	Glu	Ala	Ser	Leu	Glu	305	310	315
Glu	Gln	Gly	Leu	Ala	Val	Asp	Val	Gly	Glu	Thr	Gly	Pro	Ser	Pro	320	325	330
Pro	Ile	His	Ser	Lys	Pro	Leu	Thr	Pro	Ala	Gly	His	Thr	Gly	Gly	335	340	345
Pro	Glu	Pro	Arg	Pro	Pro	Ala	Arg	Val	Glu	Thr	Leu	Arg	Glu	Glu	350	355	360
Ala	Pro	Thr	Asp	Leu	Arg	Val	Phe	Glu	Leu	Asn	Ser	Asp	Ser	Gly	365	370	375
Lys	Ser	Thr	Pro	Ser	Asn	Asn	Gly	Lys	Lys	Gly	Ser	Ser	Thr	Asp	380	385	390
Ile	Ser	Glu	Asp	Trp	Glu	Lys	Asp	Phe	Asp	Leu	Asp	Met	Thr	Glu	395	400	405
Glu	Glu	Val	Gln	Met	Ala	Leu	Ser	Lys	Val	Asp	Ala	Ser	Gly	Glu	410	415	420
Leu	Glu	Asp	Val	Glu	Trp	Glu	Asp	Trp	Glu						425	430	

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 143 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: ADRETUT06

(B) CLONE: 2821925

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63 :

Met	Gly	Pro	Val	Arg	Leu	Gly	Ile	Leu	Leu	Phe	Leu	Phe	Leu	Ala	
				5					10					15	
Val	His	Glu	Ala	Trp	Ala	Gly	Met	Leu	Lys	Glu	Glu	Asp	Asp	Asp	
				20					25					30	
Thr	Glu	Arg	Leu	Pro	Ser	Lys	Cys	Glu	Val	Cys	Lys	Leu	Leu	Ser	
				35					40					45	
Thr	Glu	Leu	Gln	Ala	Glu	Leu	Ser	Arg	Thr	Gly	Arg	Ser	Arg	Glu	
				50					55					60	
Val	Leu	Glu	Leu	Gly	Gln	Val	Leu	Asp	Thr	Gly	Lys	Arg	Lys	Arg	
				65					70					75	
His	Val	Pro	Tyr	Ser	Val	Ser	Glu	Thr	Arg	Leu	Glu	Glu	Ala	Leu	
				80					85					90	
Glu	Asn	Leu	Cys	Glu	Arg	Ile	Leu	Asp	Tyr	Ser	Val	His	Ala	Glu	
				95					100					105	
Arg	Lys	Gly	Ser	Leu	Arg	Tyr	Ala	Lys	Gly	Gln	Ser	Gln	Thr	Met	
				110					115					120	
Ala	Thr	Leu	Lys	Gly	Leu	Val	Gln	Lys	Gly	Val	Lys	Val	Asp	Leu	
				125					130					135	
Gly	Ile	Pro	Leu	Glu	Leu	Leu	Gly								
				140											

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 301 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: UTRSTUT05

(B) CLONE: 2879068

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64 :

Met	Glu	Asp	Met	Asn	Glu	Tyr	Ser	Asn	Ile	Glu	Glu	Phe	Ala	Glu	
				5					10					15	
Gly	Ser	Lys	Ile	Asn	Ala	Ser	Lys	Asn	Gln	Gln	Asp	Asp	Gly	Lys	
				20					25					30	
Met	Phe	Ile	Gly	Gly	Leu	Ser	Trp	Asp	Thr	Ser	Lys	Lys	Asp	Leu	
				35					40					45	
Thr	Glu	Tyr	Leu	Ser	Arg	Phe	Gly	Glu	Val	Val	Asp	Cys	Thr	Ile	
				50					55					60	
Lys	Thr	Asp	Pro	Val	Thr	Gly	Arg	Ser	Arg	Gly	Phe	Gly	Phe	Val	
				65					70					75	
Leu	Phe	Lys	Asp	Ala	Ala	Ser	Val	Asp	Lys	Val	Leu	Glu	Leu	Lys	
				80					85					90	
Glu	His	Lys	Leu	Asp	Gly	Lys	Leu	Ile	Asp	Pro	Lys	Arg	Ala	Lys	
				95					100					105	
Ala	Leu	Lys	Gly	Lys	Glu	Pro	Pro	Lys	Lys	Val	Phe	Val	Gly	Gly	
				110					115					120	
Leu	Ser	Pro	Asp	Thr	Ser	Glu	Glu	Gln	Ile	Lys	Glu	Tyr	Phe	Gly	

Ala	Phe	Gly	Glu	Ile	Glu	Asn	Ile	Glu	Leu	Pro	Met	Asp	Thr	Lys
Thr	Asn	Glu	Arg	Arg	Gly	Phe	Cys	Phe	Ile	Thr	Tyr	Thr	Asp	Glu
Glu	Pro	Val	Lys	Lys	Leu	Leu	Glu	Ser	Arg	Tyr	His	Gln	Ile	Gly
Ser	Gly	Lys	Cys	Glu	Ile	Lys	Val	Ala	Gln	Pro	Lys	Glu	Val	Tyr
Arg	Gln	Gln	Gln	Gln	Gln	Gln	Lys	Gly	Gly	Arg	Gly	Ala	Ala	Ala
Gly	Gly	Arg	Gly	Gly	Thr	Arg	Gly	Arg	Gly	Arg	Gly	Gln	Gly	Gln
Asn	Trp	Asn	Gln	Gly	Phe	Asn	Asn	Tyr	Tyr	Asp	Gln	Gly	Tyr	Gly
Asn	Tyr	Asn	Ser	Ala	Tyr	Gly	Gly	Asp	Gln	Asn	Tyr	Ser	Gly	Tyr
Gly	Gly	Tyr	Asp	Tyr	Thr	Gly	Tyr	Asn	Tyr	Gly	Asn	Tyr	Gly	Tyr
Gly	Gln	Gly	Tyr	Ala	Asp	Tyr	Ser	Gly	Gln	Gln	Ser	Thr	Tyr	Gly
Lys	Ala	Ser	Arg	Gly	Gly	Gly	Asn	His	Gln	Asn	Asn	Tyr	Gln	Pro
Tyr														

## (2) INFORMATION FOR SEQ ID NO: 65:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 233 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: SINJNOT02
- (B) CLONE: 2886757

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65 :

Met	Gly	Glu	Pro	Gln	Gln	Val	Ser	Ala	Leu	Pro	Pro	Pro	Pro	Met
Gln	Tyr	Ile	Lys	Glu	Tyr	Thr	Asp	Glu	Asn	Ile	Gln	Glu	Gly	Leu
Ala	Pro	Lys	Pro	Pro	Pro	Pro	Ile	Lys	Asp	Ser	Tyr	Met	Met	Phe
Gly	Asn	Gln	Phe	Gln	Cys	Asp	Asp	Leu	Ile	Ile	Arg	Pro	Leu	Glu
Ser	Gln	Gly	Ile	Glu	Arg	Leu	His	Pro	Met	Gln	Phe	Asp	His	Lys
Lys	Glu	Leu	Arg	Lys	Leu	Asn	Met	Ser	Ile	Leu	Ile	Asn	Phe	Leu
Asp	Leu	Leu	Asp	Ile	Leu	Ile	Arg	Ser	Pro	Gly	Ser	Ile	Lys	Arg
Glu	Glu	Lys	Leu	Glu	Asp	Leu	Lys	Leu	Leu	Phe	Val	His	Val	His
His	Leu	Ile	Asn	Glu	Tyr	Arg	Pro	His	Gln	Ala	Arg	Glu	Thr	Leu
Arg	Val	Met	Met	Glu	Val	Gln	Lys	Arg	Gln	Arg	Leu	Glu	Thr	Ala

PF-0459 US

Glu	Arg	Phe	Gln	Lys	His	Leu	Glu	Arg	Val	Ile	Glu	Met	Ile	Gln
				155					160					165
Asn	Cys	Leu	Ala	Ser	Leu	Pro	Asp	Asp	Leu	Pro	His	Ser	Glu	Ala
				170					175					180
Gly	Met	Arg	Val	Lys	Thr	Glu	Pro	Met	Asp	Ala	Asp	Asp	Ser	Asn
				185					190					195
Asn	Cys	Thr	Gly	Gln	Asn	Glu	His	Gln	Arg	Glu	Asn	Ser	Gly	His
				200					205					210
Arg	Arg	Asp	Gln	Ile	Ile	Glu	Lys	Asp	Ala	Ala	Leu	Cys	Val	Leu
				215					220					225
Ile	Asp	Glu	Met	Asn	Glu	Arg	Pro							
				230										

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 354 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: SCORNOT04

(B) CLONE: 2964329

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66 :

Met	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Arg	Gly	Gly	Ala	Ala
				5					10					15
Ala	Gly	Val	Glu	Ala	Arg	Ala	Arg	Asp	Pro	Pro	Pro	Ala	His	Arg
				20					25					30
Ala	His	Pro	Arg	His	Pro	Arg	Pro	Ala	Ala	Gln	Pro	Ser	Ala	Arg
				35					40					45
Arg	Met	Asp	Gly	Gly	Ser	Gly	Gly	Leu	Gly	Ser	Gly	Asp	Asn	Ala
				50					55					60
Pro	Thr	Thr	Glu	Ala	Leu	Phe	Val	Ala	Leu	Gly	Ala	Gly	Val	Thr
				65					70					75
Ala	Leu	Ser	His	Pro	Leu	Leu	Tyr	Val	Lys	Leu	Leu	Ile	Gln	Val
				80					85					90
Gly	His	Glu	Pro	Met	Pro	Pro	Thr	Leu	Gly	Thr	Asn	Val	Leu	Gly
				95					100					105
Arg	Lys	Val	Leu	Tyr	Leu	Pro	Ser	Phe	Phe	Thr	Tyr	Ala	Lys	Tyr
				110					115					120
Ile	Val	Gln	Val	Asp	Gly	Lys	Ile	Gly	Leu	Phe	Arg	Gly	Leu	Ser
				125					130					135
Pro	Arg	Leu	Met	Ser	Asn	Ala	Leu	Ser	Thr	Val	Thr	Arg	Gly	Ser
				140					145					150
Met	Lys	Lys	Val	Phe	Pro	Pro	Asp	Glu	Ile	Glu	Gln	Val	Ser	Asn
				155					160					165
Lys	Asp	Asp	Met	Lys	Thr	Ser	Leu	Lys	Lys	Val	Val	Lys	Glu	Thr
				170					175					180
Ser	Tyr	Glu	Met	Met	Met	Gln	Cys	Val	Ser	Arg	Met	Leu	Ala	His
				185					190					195
Pro	Leu	His	Val	Ile	Ser	Met	Arg	Cys	Met	Val	Gln	Phe	Val	Gly
				200					205					210
Arg	Glu	Ala	Lys	Tyr	Ser	Gly	Val	Leu	Ser	Ser	Ile	Gly	Lys	Ile
				215					220					225
Phe	Lys	Glu	Glu	Gly	Leu	Leu	Gly	Phe	Phe	Val	Gly	Leu	Ile	Pro

His Leu Leu Gly	230	235	240
Asp Val Val Phe Leu	245	250	255
Ala His Phe Ile Asn Ala Tyr Leu Val	260	265	270
Arg Ser Tyr Thr Lys Phe Val Met Gly Ile	275	280	285
Val Ser Met Leu Thr Tyr Pro Phe Leu	290	295	300
Ala Val Asn Asn Cys Gly Leu Gln Ala	305	310	315
Pro Val Phe Lys Ser Trp Ile His Cys	320	325	330
Gln Gly Gln Leu Phe Arg Gly Ser Ser	335	340	345
Ser Ser Gly Ser Cys Phe Ala Leu Glu	350		

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 235 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: SCORNOT04
- (B) CLONE: 2965248

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67 :

Met Ala Ser Thr Ile	5	Ser Ala Tyr Lys	10	Glu Lys Met Lys	15	Glu Leu	
Ser Val Leu Ser Leu	20	Ile Cys Ser Cys	25	Phe Tyr Thr Gln	30	Pro His	
Pro Asn Thr Val Tyr	35	Gln Tyr Gly Asp	40	Met Glu Val Lys	45	Gln Leu	
Asp Lys Arg Ala Ser	50	Gly Gln Ser Phe	55	Glu Val Ile Leu	60	Lys Ser	
Pro Ser Asp Leu Ser	65	Pro Glu Ser Pro	70	Met Leu Ser Ser	75	Pro Pro	
Lys Lys Lys Asp Thr	80	Ser Leu Glu Glu	85	Leu Gln Lys Arg	90	Leu Glu	
Ala Ala Glu Glu Arg	95	Arg Lys Thr Gln	100	Glu Ala Gln Val	105	Leu Lys	
Gln Leu Ala Asp Gly	110	Ala Ser Thr Ser	115	Ala Arg Cys Cys	120	Thr Arg	
Arg Trp Arg Arg Ile	125	Thr Thr Ser Ala	130	Ala Arg Arg Arg	135	Arg Ser	
Ser Thr Thr Arg Trp	140	Ser Ser Ala Arg	145	Arg Ser Ala Arg	150	His Thr	
Trp Pro His Cys Ala	155	Ser Gly Cys Ala	160	Arg Arg Ser Cys	165	Thr Arg	
Pro Arg Cys Ala Gly	170	Thr Arg Ser Ser	175	Glu Lys Arg Cys	180	Arg Ala	
Lys Gly Pro Gly Arg	185	Ala Ala Pro Ile	190	Leu Arg Arg Asn	195	Thr Phe	



PF-0459 US

Gly	Phe	Trp	Phe	Cys	Phe	Val	His	Leu	Cys	Leu	Asp	Ala	Thr	Phe
				200					205					210
Val	Pro	Pro	Pro	Pro	Pro	Gln	Pro	Pro	Ala	Ser	Cys	Phe	Ser	Ser
				215					220					225
Ala	Leu	Ser	Arg	Pro	Ala	Leu	Ser	Ser	Trp					
				230					235					

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 221 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: TLYMNOT06
- (B) CLONE: 3000534

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68 :

Met	Trp	Ser	Ala	Gly	Arg	Gly	Gly	Ala	Ala	Trp	Pro	Val	Leu	Leu
				5					10					15
Gly	Leu	Leu	Leu	Ala	Leu	Leu	Val	Pro	Gly	Gly	Gly	Ala	Ala	Lys
				20					25					30
Thr	Gly	Ala	Glu	Leu	Val	Thr	Cys	Gly	Ser	Val	Leu	Lys	Leu	Leu
				35					40					45
Asn	Thr	His	His	Arg	Val	Arg	Leu	His	Ser	His	Asp	Ile	Lys	Tyr
				50					55					60
Gly	Ser	Gly	Ser	Gly	Gln	Gln	Ser	Val	Thr	Gly	Val	Glu	Ala	Ser
				65					70					75
Asp	Asp	Ala	Asn	Ser	Tyr	Trp	Arg	Ile	Arg	Gly	Gly	Ser	Glu	Gly
				80					85					90
Gly	Cys	Pro	Arg	Gly	Ser	Pro	Val	Arg	Cys	Gly	Gln	Ala	Val	Arg
				95					100					105
Leu	Thr	His	Val	Leu	Thr	Gly	Lys	Asn	Leu	His	Thr	His	His	Phe
				110					115					120
Pro	Ser	Pro	Leu	Ser	Asn	Asn	Gln	Glu	Val	Ser	Ala	Phe	Gly	Glu
				125					130					135
Asp	Gly	Glu	Gly	Asp	Asp	Leu	Asp	Leu	Trp	Thr	Val	Arg	Cys	Ser
				140					145					150
Gly	Gln	His	Trp	Glu	Arg	Glu	Ala	Ala	Val	Arg	Phe	Gln	His	Val
				155					160					165
Gly	Thr	Ser	Val	Phe	Leu	Ser	Val	Thr	Gly	Glu	Gln	Tyr	Gly	Ser
				170					175					180
Pro	Ile	Arg	Gly	Gln	His	Glu	Val	His	Gly	Met	Pro	Ser	Ala	Asn
				185					190					195
Thr	His	Asn	Thr	Trp	Lys	Ala	Met	Glu	Gly	Ile	Phe	Ile	Lys	Pro
				200					205					210
Ser	Val	Glu	Pro	Ser	Ala	Gly	His	Asp	Glu	Leu				
				215					220					

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 483 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

(A) LIBRARY: HEAANOT01  
 (B) CLONE: 3046870

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69 :

Met	Lys	Ala	Phe	His	Thr	Phe	Cys	Val	Val	Leu	Leu	Val	Phe	Gly	
				5					10						15
Ser	Val	Ser	Glu	Ala	Lys	Phe	Asp	Asp	Phe	Glu	Asp	Glu	Glu	Asp	
				20					25						30
Ile	Val	Glu	Tyr	Asp	Asp	Asn	Asp	Phe	Ala	Glu	Phe	Glu	Asp	Val	
				35					40						45
Met	Glu	Asp	Ser	Val	Thr	Glu	Ser	Pro	Gln	Arg	Val	Ile	Ile	Thr	
				50					55						60
Glu	Asp	Asp	Glu	Asp	Glu	Thr	Thr	Val	Glu	Leu	Glu	Gly	Gln	Asp	
				65					70						75
Glu	Asn	Gln	Glu	Gly	Asp	Phe	Glu	Asp	Ala	Asp	Thr	Gln	Glu	Gly	
				80					85						90
Asp	Thr	Glu	Ser	Glu	Pro	Tyr	Asp	Asp	Glu	Glu	Phe	Glu	Gly	Tyr	
				95					100						105
Glu	Asp	Lys	Pro	Asp	Thr	Ser	Ser	Ser	Lys	Asn	Lys	Asp	Pro	Ile	
				110					115						120
Thr	Ile	Val	Asp	Val	Pro	Ala	His	Leu	Gln	Asn	Ser	Trp	Glu	Ser	
				125					130						135
Tyr	Tyr	Leu	Glu	Ile	Leu	Met	Val	Thr	Gly	Leu	Leu	Ala	Tyr	Ile	
				140					145						150
Met	Asn	Tyr	Ile	Ile	Gly	Lys	Asn	Lys	Asn	Ser	Arg	Leu	Ala	Gln	
				155					160						165
Ala	Trp	Phe	Asn	Thr	His	Arg	Glu	Leu	Leu	Glu	Ser	Asn	Phe	Thr	
				170					175						180
Leu	Val	Gly	Asp	Asp	Gly	Thr	Asn	Lys	Glu	Ala	Thr	Ser	Thr	Gly	
				185					190						195
Lys	Leu	Asn	Gln	Glu	Asn	Glu	His	Ile	Tyr	Asn	Leu	Trp	Cys	Ser	
				200					205						210
Gly	Arg	Val	Cys	Cys	Glu	Gly	Met	Leu	Ile	Gln	Leu	Arg	Phe	Leu	
				215					220						225
Lys	Arg	Gln	Asp	Leu	Leu	Asn	Val	Leu	Ala	Arg	Met	Met	Arg	Pro	
				230					235						240
Val	Ser	Asp	Gln	Val	Gln	Ile	Lys	Val	Thr	Met	Asn	Asp	Glu	Asp	
				245					250						255
Met	Asp	Thr	Tyr	Val	Phe	Ala	Val	Gly	Thr	Arg	Lys	Ala	Leu	Val	
				260					265						270
Arg	Leu	Gln	Lys	Glu	Met	Gln	Asp	Leu	Ser	Glu	Phe	Cys	Ser	Asp	
				275					280						285
Lys	Pro	Lys	Ser	Gly	Ala	Lys	Tyr	Gly	Leu	Pro	Asp	Ser	Leu	Ala	
				290					295						300
Ile	Leu	Ser	Glu	Met	Gly	Glu	Val	Thr	Asp	Gly	Met	Met	Asp	Thr	
				305					310						315
Lys	Met	Val	His	Phe	Leu	Thr	His	Tyr	Ala	Asp	Lys	Ile	Glu	Ser	
				320					325						330
Val	His	Phe	Ser	Asp	Gln	Phe	Ser	Gly	Pro	Lys	Ile	Met	Gln	Glu	
				335					340						345
Glu	Gly	Gln	Pro	Leu	Lys	Leu	Pro	Asp	Thr	Lys	Arg	Thr	Leu	Leu	
				350					355						360
Phe	Thr	Phe	Asn	Val	Pro	Gly	Ser	Gly	Asn	Thr	Tyr	Pro	Lys	Asp	
				365					370						375

Met	Glu	Ala	Leu	Leu	Pro	Leu	Met	Asn	Met	Val	Ile	Tyr	Ser	Ile		
				380					385						390	
Asp	Lys	Ala	Lys	Lys	Phe	Arg	Leu	Asn	Arg	Glu	Gly	Lys	Gln	Lys		
				395					400						405	
Ala	Asp	Lys	Asn	Arg	Ala	Arg	Val	Glu	Glu	Asn	Phe	Leu	Lys	Leu		
				410					415						420	
Thr	His	Val	Gln	Arg	Gln	Glu	Ala	Ala	Gln	Ser	Arg	Arg	Glu	Glu		
				425					430						435	
Lys	Lys	Arg	Ala	Glu	Lys	Glu	Arg	Ile	Met	Asn	Glu	Glu	Asp	Pro		
				440					445						450	
Glu	Lys	Gln	Arg	Arg	Leu	Glu	Glu	Ala	Ala	Leu	Arg	Arg	Glu	Gln		
				455					460						465	
Lys	Lys	Leu	Glu	Lys	Lys	Gln	Met	Lys	Met	Lys	Gln	Ile	Lys	Val		
				470					475						480	
Lys	Ala	Met														

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 371 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PONSAT01
- (B) CLONE: 3057669

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70 :

Met	Asp	His	Glu	Asp	Ile	Ser	Glu	Ser	Val	Asp	Ala	Ala	Tyr	Asn		
				5					10					15		
Leu	Gln	Asp	Ser	Cys	Leu	Thr	Asp	Cys	Asp	Val	Glu	Asp	Gly	Thr		
				20					25					30		
Met	Asp	Gly	Asn	Asp	Glu	Gly	His	Ser	Phe	Glu	Leu	Cys	Pro	Ser		
				35					40					45		
Glu	Ala	Ser	Pro	Tyr	Val	Arg	Ser	Arg	Glu	Arg	Thr	Ser	Ser	Ser		
				50					55					60		
Ile	Val	Phe	Glu	Asp	Ser	Gly	Cys	Asp	Asn	Ala	Ser	Ser	Lys	Glu		
				65					70					75		
Glu	Pro	Lys	Thr	Asn	Arg	Leu	His	Ile	Gly	Asn	His	Cys	Ala	Asn		
				80					85					90		
Lys	Leu	Thr	Ala	Phe	Lys	Pro	Thr	Ser	Ser	Lys	Ser	Ser	Ser	Glu		
				95					100					105		
Ala	Thr	Leu	Ser	Ile	Ser	Pro	Pro	Arg	Pro	Thr	Thr	Leu	Ser	Leu		
				110					115					120		
Asp	Leu	Thr	Lys	Asn	Thr	Thr	Glu	Lys	Leu	Gln	Pro	Ser	Ser	Pro		
				125					130					135		
Lys	Val	Tyr	Leu	Tyr	Ile	Gln	Met	Gln	Leu	Cys	Arg	Lys	Glu	Asn		
				140					145					150		
Leu	Lys	Asp	Trp	Met	Asn	Gly	Arg	Cys	Thr	Ile	Glu	Glu	Arg	Glu		
				155					160					165		
Arg	Ser	Val	Cys	Leu	His	Ile	Phe	Leu	Gln	Ile	Ala	Glu	Ala	Val		
				170					175					180		
Glu	Phe	Leu	His	Ser	Lys	Gly	Leu	Met	His	Arg	Asp	Leu	Lys	Pro		
				185					190					195		
Ser	Asn	Ile	Phe	Phe	Thr	Met	Asp	Asp	Val	Val	Lys	Val	Gly	Asp		
				200					205					210		
Phe	Gly	Leu	Val	Thr	Ala	Met	Asp	Gln	Asp	Glu	Glu	Glu	Gln	Thr		

Val	Leu	Thr	Pro	Met	Pro	Ala	Tyr	Ala	Arg	His	Thr	Gly	Gln	Val	215	220	225
Gly	Thr	Lys	Leu	Tyr	Met	Ser	Pro	Glu	Gln	Ile	His	Gly	Asn	Ser	230	235	240
Tyr	Ser	His	Lys	Val	Asp	Ile	Phe	Ser	Leu	Gly	Leu	Ile	Leu	Phe	245	250	255
Glu	Leu	Leu	Tyr	Pro	Phe	Ser	Thr	Gln	Met	Glu	Arg	Val	Arg	Thr	260	265	270
Leu	Thr	Asp	Val	Arg	Asn	Leu	Lys	Phe	Pro	Pro	Leu	Phe	Thr	Gln	275	280	285
Lys	Tyr	Pro	Cys	Glu	Tyr	Val	Met	Val	Gln	Asp	Met	Leu	Ser	Pro	290	295	300
Ser	Pro	Met	Glu	Arg	Pro	Glu	Ala	Ile	Asn	Ile	Ile	Glu	Asn	Ala	305	310	315
Val	Phe	Glu	Asp	Leu	Asp	Phe	Pro	Gly	Lys	Thr	Val	Leu	Arg	Gln	320	325	330
Arg	Ser	Arg	Ser	Leu	Ser	Ser	Ser	Gly	Thr	Lys	His	Ser	Arg	Gln	335	340	345
Ser	Asn	Asn	Ser	His	Ser	Pro	Leu	Pro	Ser	Asn					350	355	360
															365	370	

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 402 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: HEAONOT03
- (B) CLONE: 3088178

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71 :

Met	Met	Asn	Asn	Arg	Phe	Arg	Lys	Asp	Met	Met	Lys	Asn	Ala	Ser	15
Glu	Ser	Lys	Leu	Ser	Lys	Asp	Asn	Leu	Lys	Lys	Arg	Leu	Lys	Glu	30
Glu	Phe	Gln	His	Ala	Met	Gly	Gly	Val	Pro	Ala	Trp	Ala	Glu	Thr	45
Thr	Lys	Arg	Lys	Thr	Ser	Ser	Asp	Asp	Glu	Ser	Glu	Glu	Asp	Glu	60
Asp	Asp	Leu	Leu	Gln	Arg	Thr	Gly	Asn	Phe	Ile	Ser	Thr	Ser	Thr	75
Ser	Leu	Pro	Arg	Gly	Ile	Leu	Lys	Met	Lys	Asn	Cys	Gln	His	Ala	90
Asn	Ala	Glu	Arg	Pro	Thr	Val	Ala	Arg	Ile	Ser	Ser	Val	Gln	Phe	105
His	Pro	Gly	Ala	Gln	Ile	Val	Met	Val	Ala	Gly	Leu	Asp	Asn	Ala	120
Val	Ser	Leu	Phe	Gln	Val	Asp	Gly	Lys	Thr	Asn	Pro	Lys	Ile	Gln	135
Ser	Ile	Tyr	Leu	Glu	Arg	Phe	Pro	Ile	Phe	Lys	Ala	Cys	Phe	Ser	150
Ala	Asn	Gly	Glu	Glu	Val	Leu	Ala	Thr	Ser	Thr	His	Ser	Lys	Val	165

Leu Tyr Val Tyr	Asp Met Leu Ala Gly	Lys Leu Ile Pro Val	His
	170	175	180
Gln Val Arg Gly	Leu Lys Glu Lys Ile	Val Arg Ser Phe Glu	Val
	185	190	195
Ser Pro Asp Gly	Ser Phe Leu Leu Ile	Asn Gly Ile Ala Gly	Tyr
	200	205	210
Leu His Leu Leu	Ala Met Lys Thr Lys	Glu Leu Ile Gly Ser	Met
	215	220	225
Lys Ile Asn Gly	Arg Val Ala Ala Ser	Thr Phe Ser Ser Asp	Ser
	230	235	240
Lys Lys Val Tyr	Ala Ser Ser Gly Asp	Gly Glu Val Tyr Val	Trp
	245	250	255
Asp Val Asn Ser	Arg Lys Cys Leu Asn	Arg Phe Val Asp Glu	Gly
	260	265	270
Ser Leu Tyr Gly	Leu Ser Ile Ala Thr	Ser Arg Asn Gly Gln	Tyr
	275	280	285
Val Ala Cys Gly	Ser Asn Cys Gly Val	Val Asn Ile Tyr Asn	Gln
	290	295	300
Asp Ser Cys Leu	Gln Glu Thr Asn Pro	Lys Pro Ile Lys Ala	Ile
	305	310	315
Met Asn Leu Val	Thr Gly Val Thr Ser	Leu Thr Phe Asn Pro	Thr
	320	325	330
Thr Glu Ile Leu	Ala Ile Ala Ser Glu	Lys Met Lys Glu Ala	Val
	335	340	345
Arg Leu Val His	Leu Pro Ser Cys Thr	Val Phe Ser Asn Phe	Pro
	350	355	360
Val Ile Lys Asn	Lys Asn Ile Ser His	Val His Thr Met Asp	Phe
	365	370	375
Ser Pro Arg Ser	Gly Tyr Phe Ala Leu	Gly Asn Glu Lys Gly	Lys
	380	385	390
Ala Leu Met Tyr	Arg Leu His His Tyr	Ser Asp Phe	
	395	400	

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 640 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRSTNOT19
- (B) CLONE: 3094321

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72 :

Met Ala Leu Ser Arg	Gly Leu Pro Arg	Glu Leu Ala Glu Ala Val
	5	10
Ala Gly Gly Arg Val	Leu Val Val Gly	Ala Gly Gly Ile Gly Cys
	20	25
Glu Leu Leu Lys Asn	Leu Val Leu Thr	Gly Phe Ser His Ile Asp
	35	40
Leu Ile Asp Leu Asp	Thr Ile Asp Val	Ser Asn Leu Asn Arg Gln
	50	55
Phe Leu Phe Gln Lys	Lys His Val Gly	Ser Lys Ala Gln Val
	65	70
Ala Lys Glu Ser Val	Leu Gln Phe Tyr	Pro Lys Ala Asn Ile Val

	80		85		90
Ala Tyr His Asp	Ser Ile Met Asn Pro	Asp Tyr Asn Val Glu	Phe		
	95		100		105
Phe Arg Gln Phe	Ile Leu Val Met Asn	Ala Leu Asp Asn Arg	Ala		
	110		115		120
Ala Arg Asn His	Val Asn Arg Met Cys	Leu Ala Ala Asp Val	Pro		
	125		130		135
Leu Ile Glu Ser	Gly Thr Ala Gly Tyr	Leu Gly Gln Val Thr	Thr		
	140		145		150
Ile Lys Lys Gly	Val Thr Glu Cys Tyr	Glu Cys His Pro Lys	Pro		
	155		160		165
Thr Gln Arg Thr	Phe Pro Gly Cys Thr	Ile Arg Asn Thr Pro	Ser		
	170		175		180
Glu Pro Ile His	Cys Ile Val Trp Ala	Lys Tyr Leu Phe Asn	Gln		
	185		190		195
Leu Phe Gly Glu	Glu Asp Ala Asp Gln	Glu Val Ser Pro Asp	Arg		
	200		205		210
Ala Asp Pro Glu	Ala Ala Trp Glu Pro	Thr Glu Ala Glu Ala	Arg		
	215		220		225
Ala Arg Ala Ser	Asn Glu Asp Gly Asp	Ile Lys Arg Ile Ser	Thr		
	230		235		240
Lys Glu Trp Ala	Lys Ser Thr Gly Tyr	Asp Pro Val Lys Leu	Phe		
	245		250		255
Thr Lys Leu Phe	Lys Asp Asp Ile Arg	Tyr Leu Leu Thr Met	Asp		
	260		265		270
Lys Leu Trp Arg	Lys Arg Lys Pro Pro	Val Pro Leu Asp Trp	Ala		
	275		280		285
Glu Val Gln Ser	Gln Gly Glu Glu Thr	Asn Ala Ser Asp Gln	Gln		
	290		295		300
Asn Glu Pro Gln	Leu Gly Leu Lys Asp	Gln Gln Val Leu Asp	Val		
	305		310		315
Lys Ser Tyr Ala	Arg Leu Phe Ser Lys	Ser Ile Glu Thr Leu	Arg		
	320		325		330
Val His Leu Ala	Glu Lys Gly Asp Gly	Ala Glu Leu Ile Trp	Asp		
	335		340		345
Lys Asp Asp Pro	Ser Ala Met Asp Phe	Val Thr Ser Ala Ala	Asn		
	350		355		360
Leu Arg Met His	Ile Phe Ser Met Asn	Met Lys Ser Arg Phe	Asp		
	365		370		375
Ile Lys Ser Met	Ala Gly Asn Ile Ile	Pro Ala Ile Ala Thr	Thr		
	380		385		390
Asn Ala Val Ile	Ala Gly Leu Ile Val	Leu Glu Gly Leu Lys	Ile		
	395		400		405
Leu Ser Gly Lys	Ile Asp Gln Cys Arg	Thr Ile Phe Leu Asn	Lys		
	410		415		420
Gln Pro Asn Pro	Arg Lys Lys Leu Leu	Val Pro Cys Ala Leu	Asp		
	425		430		435
Pro Pro Asn Pro	Asn Cys Tyr Val Cys	Ala Ser Lys Pro Glu	Val		
	440		445		450
Thr Val Arg Leu	Asn Val His Lys Val	Thr Val Leu Thr Leu	Gln		
	455		460		465
Asp Lys Ile Val	Lys Glu Lys Phe Ala	Met Val Ala Pro Asp	Val		
	470		475		480
Gln Ile Glu Asp	Gly Lys Gly Thr Ile	Leu Ile Ser Ser Glu	Glu		
	485		490		495
Gly Glu Thr Glu	Ala Asn Asn His Lys	Lys Leu Ser Glu Phe	Gly		
	500		505		510
Ile Arg Asn Gly	Ser Arg Leu Gln Ala	Asp Asp Phe Leu Gln	Asp		
	515		520		525
Tyr Thr Leu Leu	Ile Asn Ile Leu His	Ser Glu Asp Leu Gly	Lys		
	530		535		540

Asp	Val	Glu	Phe	Glu	Val	Val	Gly	Asp	Ala	Pro	Glu	Lys	Val	Gly	
				545					550					555	
Pro	Lys	Gln	Ala	Glu	Asp	Ala	Ala	Lys	Ser	Ile	Thr	Asn	Gly	Ser	
				560					565					570	
Asp	Asp	Gly	Ala	Gln	Pro	Ser	Thr	Ser	Thr	Ala	Gln	Glu	Gln	Asp	
				575					580					585	
Asp	Val	Leu	Ile	Val	Asp	Ser	Asp	Glu	Glu	Asp	Ser	Ser	Asn	Asn	
				590					595					600	
Ala	Asp	Val	Ser	Glu	Glu	Glu	Arg	Ser	Arg	Lys	Arg	Lys	Leu	Asp	
				605					610					615	
Glu	Lys	Glu	Asn	Leu	Ser	Ala	Lys	Arg	Ser	Arg	Ile	Glu	Gln	Lys	
				620					625					630	
Glu	Glu	Leu	Asp	Asp	Val	Ile	Ala	Leu	Asp						
				635					640						

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 237 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGTUT13
- (B) CLONE: 3115936

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73 :

Met	Asp	Lys	Ile	Leu	Asn	Val	Glu	Glu	Thr	Tyr	Leu	Thr	Val	Leu	
				5					10					15	
Val	Lys	Ile	Gly	Pro	Gly	Phe	His	Thr	Arg	Glu	Cys	Phe	Leu	Leu	
				20					25					30	
Lys	Ser	Ile	Leu	Cys	Phe	Ser	Pro	Ser	Tyr	Arg	Met	Ser	Glu	Gly	
				35					40					45	
Asp	Ser	Val	Gly	Glu	Ser	Val	His	Gly	Lys	Pro	Ser	Val	Val	Tyr	
				50					55					60	
Arg	Phe	Phe	Thr	Arg	Leu	Gly	Gln	Ile	Tyr	Gln	Ser	Trp	Leu	Asp	
				65					70					75	
Lys	Ser	Thr	Pro	Tyr	Thr	Ala	Val	Arg	Trp	Val	Val	Thr	Leu	Gly	
				80					85					90	
Leu	Ser	Phe	Val	Tyr	Met	Ile	Arg	Val	Tyr	Leu	Leu	Gln	Gly	Trp	
				95					100					105	
Tyr	Ile	Val	Thr	Tyr	Ala	Leu	Gly	Ile	Tyr	His	Leu	Asn	Leu	Phe	
				110					115					120	
Ile	Ala	Phe	Leu	Ser	Pro	Lys	Val	Asp	Pro	Ser	Leu	Met	Glu	Asp	
				125					130					135	
Ser	Asp	Asp	Gly	Pro	Ser	Leu	Pro	Thr	Lys	Gln	Asn	Glu	Glu	Phe	
				140					145					150	
Arg	Pro	Phe	Ile	Arg	Arg	Leu	Pro	Glu	Phe	Lys	Phe	Trp	His	Ala	
				155					160					165	
Ala	Thr	Lys	Gly	Ile	Leu	Val	Ala	Met	Val	Cys	Thr	Phe	Phe	Asp	
				170					175					180	
Ala	Phe	Asn	Val	Pro	Val	Phe	Trp	Pro	Ile	Leu	Val	Met	Tyr	Phe	
				185					190					195	
Ile	Met	Leu	Phe	Cys	Ile	Thr	Met	Lys	Arg	Gln	Ile	Lys	His	Met	
				200					205					210	
Ile	Lys	Tyr	Arg	Tyr	Ile	Pro	Phe	Thr	His	Gly	Lys	Arg	Arg	Tyr	

	215	220	225
Arg Gly Lys Glu Asp	Ala Gly Lys Ala Phe	Ala Ser	
	230	235	

(2) INFORMATION FOR SEQ ID NO: 74:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 432 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGTUT13
- (B) CLONE: 3116522

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74 :

Met	Asp	Ala	Arg	Trp	Trp	Ala	Val	Val	Val	Leu	Ala	Ala	Phe	Pro
				5					10					15
Ser	Leu	Gly	Ala	Gly	Gly	Glu	Thr	Pro	Glu	Ala	Pro	Pro	Glu	Ser
				20					25					30
Trp	Thr	Gln	Leu	Trp	Phe	Phe	Arg	Phe	Val	Val	Asn	Ala	Ala	Gly
				35					40					45
Tyr	Ala	Ser	Phe	Met	Val	Pro	Gly	Tyr	Leu	Leu	Val	Gln	Tyr	Phe
				50					55					60
Arg	Arg	Lys	Asn	Tyr	Leu	Glu	Thr	Gly	Arg	Gly	Leu	Cys	Phe	Pro
				65					70					75
Leu	Val	Lys	Ala	Cys	Val	Phe	Gly	Asn	Glu	Pro	Lys	Ala	Ser	Asp
				80					85					90
Glu	Val	Pro	Leu	Ala	Pro	Arg	Thr	Glu	Ala	Ala	Glu	Thr	Thr	Pro
				95					100					105
Met	Trp	Gln	Ala	Leu	Lys	Leu	Leu	Phe	Cys	Ala	Thr	Gly	Leu	Gln
				110					115					120
Val	Ser	Tyr	Leu	Thr	Trp	Gly	Val	Leu	Gln	Glu	Arg	Val	Met	Thr
				125					130					135
Arg	Ser	Tyr	Gly	Ala	Thr	Ala	Thr	Ser	Pro	Gly	Glu	Arg	Phe	Thr
				140					145					150
Asp	Ser	Gln	Phe	Leu	Val	Leu	Met	Asn	Arg	Val	Leu	Ala	Leu	Ile
				155					160					165
Val	Ala	Gly	Leu	Ser	Cys	Val	Leu	Cys	Lys	Gln	Pro	Arg	His	Gly
				170					175					180
Ala	Pro	Met	Tyr	Arg	Tyr	Ser	Phe	Ala	Ser	Leu	Ser	Asn	Val	Leu
				185					190					195
Ser	Ser	Trp	Cys	Gln	Tyr	Glu	Ala	Leu	Lys	Phe	Val	Ser	Phe	Pro
				200					205					210
Thr	Gln	Val	Leu	Ala	Lys	Ala	Ser	Lys	Val	Ile	Pro	Val	Met	Leu
				215					220					225
Met	Gly	Lys	Leu	Val	Ser	Arg	Arg	Ser	Tyr	Glu	His	Trp	Glu	Tyr
				230					235					240
Leu	Thr	Ala	Thr	Leu	Ile	Ser	Ile	Gly	Val	Ser	Met	Phe	Leu	Leu
				245					250					255
Ser	Ser	Gly	Pro	Glu	Pro	Arg	Ser	Ser	Pro	Ala	Thr	Thr	Leu	Ser
				260					265					270
Gly	Leu	Ile	Leu	Leu	Ala	Gly	Tyr	Ile	Ala	Phe	Asp	Ser	Phe	Thr
				275					280					285
Ser	Asn	Trp	Gln	Asp	Ala	Leu	Phe	Ala	Tyr	Lys	Met	Ser	Ser	Val
				290					295					300



Gln	Met	Met	Phe	Gly	Val	Asn	Phe	Phe	Ser	Cys	Leu	Phe	Thr	Val	
				305					310					315	
Gly	Ser	Leu	Leu	Glu	Gln	Gly	Ala	Leu	Leu	Glu	Gly	Thr	Arg	Phe	
				320					325					330	
Met	Gly	Arg	His	Ser	Glu	Phe	Ala	Ala	His	Ala	Leu	Leu	Leu	Ser	
				335					340					345	
Ile	Cys	Ser	Ala	Cys	Gly	Gln	Leu	Phe	Ile	Phe	Tyr	Thr	Ile	Gly	
				350					355					360	
Gln	Phe	Gly	Ala	Ala	Val	Phe	Thr	Ile	Ile	Met	Thr	Leu	Arg	Gln	
				365					370					375	
Ala	Phe	Ala	Ile	Leu	Leu	Ser	Cys	Leu	Leu	Tyr	Gly	His	Thr	Val	
				380					385					390	
Thr	Val	Val	Gly	Gly	Leu	Gly	Val	Ala	Val	Val	Phe	Ala	Ala	Leu	
				395					400					405	
Leu	Leu	Arg	Val	Tyr	Ala	Arg	Gly	Arg	Leu	Lys	Gln	Arg	Gly	Lys	
				410					415					420	
Lys	Ala	Val	Pro	Val	Glu	Ser	Pro	Val	Gln	Lys	Val				
				425					430						

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 252 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGTUT13
- (B) CLONE: 3117184

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75 :

Met	Ser	Phe	Pro	Pro	His	Leu	Asn	Arg	Pro	Pro	Met	Gly	Ile	Pro	
				5					10					15	
Ala	Leu	Pro	Pro	Gly	Thr	Pro	Pro	Pro	Gln	Phe	Pro	Gly	Phe	Pro	
				20					25					30	
Pro	Pro	Val	Pro	Pro	Gly	Thr	Pro	Met	Ile	Pro	Val	Pro	Met	Ser	
				35					40					45	
Ile	Met	Ala	Pro	Ala	Pro	Thr	Val	Leu	Val	Pro	Thr	Val	Ser	Met	
				50					55					60	
Val	Gly	Lys	His	Leu	Gly	Ala	Arg	Lys	Asp	His	Pro	Gly	Leu	Lys	
				65					70					75	
Ala	Lys	Glu	Asn	Asp	Glu	Asn	Cys	Gly	Pro	Thr	Thr	Thr	Val	Phe	
				80					85					90	
Val	Gly	Asn	Ile	Ser	Glu	Lys	Ala	Ser	Asp	Met	Leu	Ile	Arg	Gln	
				95					100					105	
Leu	Leu	Ala	Lys	Cys	Gly	Leu	Val	Leu	Ser	Trp	Lys	Arg	Val	Gln	
				110					115					120	
Gly	Ala	Ser	Gly	Lys	Leu	Gln	Ala	Phe	Gly	Phe	Cys	Glu	Tyr	Lys	
				125					130					135	
Glu	Pro	Glu	Ser	Thr	Leu	Arg	Ala	Leu	Arg	Leu	Leu	His	Asp	Leu	
				140					145					150	
Gln	Ile	Gly	Glu	Lys	Lys	Leu	Leu	Val	Lys	Val	Asp	Ala	Lys	Thr	
				155					160					165	
Lys	Ala	Gln	Leu	Asp	Glu	Trp	Lys	Ala	Lys	Lys	Lys	Ala	Ser	Asn	
				170					175					180	
Gly	Asn	Ala	Arg	Pro	Glu	Thr	Val	Thr	Asn	Asp	Asp	Glu	Glu	Ala	

Leu	Asp	Glu	Glu	185	Thr	Lys	Arg	Arg	Asp	190	Gln	Met	Ile	Lys	Gly	195	Ala
				200						205							210
Ile	Glu	Val	Leu	215	Ile	Arg	Glu	Tyr	Ser	220	Ser	Glu	Leu	Asn	Ala	Pro	225
Ser	Gln	Glu	Ser	230	Asp	Ser	His	Pro	Arg	235	Lys	Lys	Lys	Lys	Glu	Lys	240
Lys	Glu	Asp	Ile	245	Phe	Gly	Arg	Phe	Gln	250	Trp	Ala	His				

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 523 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LNODNOT05
- (B) CLONE: 3125156

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76 :

Met	Gly	Pro	Gln	Ala	Ala	Pro	Leu	Thr	Ile	Arg	Gly	Pro	Ser	Ser			
				5					10					15			
Ala	Gly	Gln	Ser	Thr	Pro	Ser	Pro	His	Leu	Val	Pro	Ser	Pro	Ala			
				20					25					30			
Pro	Ser	Pro	Gly	Pro	Gly	Pro	Val	Pro	Pro	Arg	Pro	Pro	Ala	Ala			
				35					40					45			
Glu	Pro	Pro	Pro	Cys	Leu	Arg	Arg	Gly	Ala	Ala	Ala	Ala	Asp	Leu			
				50					55					60			
Leu	Ser	Ser	Ser	Pro	Glu	Ser	Gln	His	Gly	Gly	Thr	Gln	Ser	Pro			
				65					70					75			
Gly	Gly	Gly	Gln	Pro	Leu	Leu	Gln	Pro	Thr	Lys	Val	Asp	Ala	Ala			
				80					85					90			
Glu	Gly	Arg	Arg	Pro	Gln	Ala	Leu	Arg	Leu	Ile	Glu	Arg	Asp	Pro			
				95					100					105			
Tyr	Glu	His	Pro	Glu	Arg	Leu	Arg	Gln	Leu	Gln	Gln	Glu	Leu	Glu			
				110					115					120			
Ala	Phe	Arg	Gly	Gln	Leu	Gly	Asp	Val	Gly	Ala	Leu	Asp	Thr	Val			
				125					130					135			
Trp	Arg	Glu	Leu	Gln	Asp	Ala	Gln	Glu	His	Asp	Ala	Arg	Gly	Arg			
				140					145					150			
Ser	Ile	Ala	Ile	Ala	Arg	Cys	Tyr	Ser	Leu	Lys	Asn	Arg	His	Gln			
				155					160					165			
Asp	Val	Met	Pro	Tyr	Asp	Ser	Asn	Arg	Val	Val	Leu	Arg	Ser	Gly			
				170					175					180			
Lys	Asp	Asp	Tyr	Ile	Asn	Ala	Ser	Cys	Val	Glu	Gly	Leu	Ser	Pro			
				185					190					195			
Tyr	Cys	Pro	Pro	Leu	Val	Ala	Thr	Gln	Ala	Pro	Leu	Pro	Gly	Thr			
				200					205					210			
Ala	Ala	Asp	Phe	Trp	Leu	Met	Val	His	Glu	Gln	Lys	Val	Ser	Val			
				215					220					225			
Ile	Val	Met	Leu	Val	Ser	Glu	Ala	Glu	Met	Glu	Lys	Gln	Lys	Val			
				230					235					240			
Ala	Arg	Tyr	Phe	Pro	Thr	Glu	Arg	Gly	Gln	Pro	Met	Val	His	Gly			
				245					250					255			

PF-0459 US

Ala	Leu	Ser	Leu	Ala	Leu	Ser	Ser	Val	Arg	Ser	Thr	Glu	Thr	His			
				260					265					270			
Val	Glu	Arg	Val	Leu	Ser	Leu	Gln	Phe	Arg	Asp	Gln	Ser	Leu	Lys			
				275					280					285			
Arg	Ser	Leu	Val	His	Leu	His	Phe	Pro	Thr	Trp	Pro	Glu	Leu	Gly			
				290					295					300			
Leu	Pro	Asp	Ser	Pro	Ser	Asn	Leu	Leu	Arg	Phe	Ile	Gln	Glu	Val			
				305					310					315			
His	Ala	His	Tyr	Leu	His	Gln	Arg	Pro	Leu	His	Thr	Pro	Ile	Ile			
				320					325					330			
Val	His	Cys	Ser	Ser	Gly	Val	Gly	Arg	Thr	Gly	Ala	Phe	Ala	Leu			
				335					340					345			
Leu	Tyr	Ala	Ala	Val	Gln	Glu	Val	Glu	Ala	Gly	Asn	Gly	Ile	Pro			
				350					355					360			
Glu	Leu	Pro	Gln	Leu	Val	Arg	Arg	Met	Arg	Gln	Gln	Arg	Lys	His			
				365					370					375			
Met	Leu	Gln	Glu	Lys	Leu	His	Leu	Arg	Phe	Cys	Tyr	Glu	Ala	Val			
				380					385					390			
Val	Arg	His	Val	Glu	Gln	Val	Leu	Gln	Arg	His	Gly	Val	Pro	Pro			
				395					400					405			
Pro	Cys	Lys	Pro	Leu	Ala	Ser	Ala	Ser	Ile	Ser	Gln	Lys	Asn	His			
				410					415					420			
Leu	Pro	Gln	Asp	Ser	Gln	Asp	Leu	Val	Leu	Gly	Gly	Asp	Val	Pro			
				425					430					435			
Ile	Ser	Ser	Ile	Gln	Ala	Thr	Ile	Ala	Lys	Leu	Ser	Ile	Arg	Pro			
				440					445					450			
Pro	Gly	Gly	Leu	Glu	Ser	Pro	Val	Ala	Ser	Leu	Pro	Gly	Pro	Ala			
				455					460					465			
Glu	Pro	Pro	Gly	Leu	Pro	Pro	Ala	Ser	Leu	Pro	Glu	Ser	Thr	Pro			
				470					475					480			
Ile	Pro	Ser	Ser	Ser	Gln	Thr	Pro	Phe	Pro	Pro	His	Tyr	Leu	Arg			
				485					490					495			
Leu	Pro	Ser	Leu	Arg	Arg	Ser	Arg	Gln	Cys	Leu	Lys	Pro	Pro	Ala			
				500					505					510			
Arg	Gly	Pro	Pro	Pro	Pro	Pro	Trp	Asn	Cys	Trp	Pro	Pro					
				515					520								

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 621 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGTUT12
- (B) CLONE: 3129120

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77 :

Met	Gly	Leu	Leu	Ser	Asp	Pro	Val	Arg	Arg	Arg	Ala	Leu	Ala	Arg			
				5					10					15			
Leu	Val	Leu	Arg	Leu	Asn	Ala	Pro	Leu	Cys	Val	Leu	Ser	Tyr	Val			
				20					25					30			
Ala	Gly	Ile	Ala	Trp	Phe	Leu	Ala	Leu	Val	Phe	Pro	Pro	Leu	Thr			
				35					40					45			
Gln	Arg	Thr	Tyr	Met	Ser	Glu	Asn	Ala	Met	Gly	Ser	Thr	Met	Val			
				50					55					60			

Glu Glu Gln Phe Ala Gly Gly Asp Arg Ala Arg Ala Phe Ala Arg  
 65 70 75  
 Asp Phe Ala Ala His Arg Lys Lys Ser Gly Ala Leu Pro Val Ala  
 80 85 90  
 Trp Leu Glu Arg Thr Met Arg Ser Val Gly Leu Glu Val Tyr Thr  
 95 100 105  
 Gln Ser Phe Ser Arg Lys Leu Pro Phe Pro Asp Glu Thr His Glu  
 110 115 120  
 Arg Tyr Met Val Ser Gly Thr Asn Val Tyr Gly Ile Leu Arg Ala  
 125 130 135  
 Pro Arg Ala Ala Ser Thr Glu Ser Leu Val Leu Thr Val Pro Cys  
 140 145 150  
 Gly Ser Asp Ser Thr Asn Ser Gln Ala Val Gly Leu Leu Leu Ala  
 155 160 165  
 Leu Ala Ala His Phe Arg Gly Gln Ile Tyr Trp Ala Lys Asp Ile  
 170 175 180  
 Val Phe Leu Val Thr Glu His Asp Leu Leu Gly Thr Glu Ala Trp  
 185 190 195  
 Leu Glu Ala Tyr His Asp Val Asn Val Thr Gly Met Gln Ser Ser  
 200 205 210  
 Pro Leu Gln Gly Arg Ala Gly Ala Ile Gln Ala Ala Val Ala Leu  
 215 220 225  
 Glu Leu Ser Ser Asp Val Val Thr Ser Leu Asp Val Ala Val Glu  
 230 235 240  
 Gly Leu Asn Gly Gln Leu Pro Asn Leu Asp Leu Leu Asn Leu Phe  
 245 250 255  
 Gln Thr Phe Cys Gln Lys Gly Gly Leu Leu Cys Thr Leu Gln Gly  
 260 265 270  
 Lys Leu Gln Pro Glu Asp Trp Thr Ser Leu Asp Gly Pro Leu Gln  
 275 280 285  
 Gly Leu Gln Thr Leu Leu Leu Met Val Leu Arg Gln Ala Ser Gly  
 290 295 300  
 Arg Pro His Gly Ser His Gly Leu Phe Leu Arg Tyr Arg Val Glu  
 305 310 315  
 Ala Leu Thr Leu Arg Gly Ile Asn Ser Phe Arg Gln Tyr Lys Tyr  
 320 325 330  
 Asp Leu Val Ala Val Gly Lys Ala Leu Glu Gly Met Phe Arg Lys  
 335 340 345  
 Leu Asn His Leu Leu Glu Arg Leu His Gln Ser Phe Phe Leu Tyr  
 350 355 360  
 Leu Leu Pro Gly Leu Ser Arg Phe Val Ser Ile Gly Leu Tyr Met  
 365 370 375  
 Pro Ala Val Gly Phe Leu Leu Leu Val Leu Gly Leu Lys Ala Leu  
 380 385 390  
 Glu Leu Trp Met Gln Leu His Glu Ala Gly Met Gly Leu Glu Glu  
 395 400 405  
 Pro Gly Gly Ala Pro Gly Pro Ser Val Pro Leu Pro Pro Ser Gln  
 410 415 420  
 Gly Val Gly Leu Ala Ser Leu Val Ala Pro Leu Leu Ile Ser Gln  
 425 430 435  
 Ala Met Gly Leu Ala Leu Tyr Val Leu Pro Val Leu Gly Gln His  
 440 445 450  
 Val Ala Thr Gln His Phe Pro Val Ala Glu Ala Glu Ala Val Val  
 455 460 465  
 Leu Thr Leu Leu Ala Ile Tyr Ala Ala Gly Leu Ala Leu Pro His  
 470 475 480  
 Asn Thr His Arg Val Val Ser Thr Gln Ala Pro Asp Arg Gly Trp  
 485 490 495  
 Met Ala Leu Lys Leu Val Ala Leu Ile Tyr Leu Ala Leu Gln Leu  
 500 505 510  
 Gly Cys Ile Ala Leu Thr Asn Phe Ser Leu Gly Phe Leu Leu Ala

Thr Thr Met Val	515	Pro Thr Ala Ala Leu	520	Ala Lys Pro His Gly	525
	530		535		540
Arg Thr Leu Tyr	545	Ala Ala Leu Leu Val	550	Leu Thr Ser Pro Ala	555
	560		565		570
Thr Leu Leu Gly	575	Ser Leu Phe Leu Trp	580	Arg Glu Leu Gln Glu	585
	590		595		600
Pro Leu Ser Leu	605	Ala Glu Gly Trp Gln	610	Leu Phe Leu Ala Ala	615
	620				
Ala Gln Gly Val		Leu Glu His His Thr		Tyr Gly Ala Leu Leu	
Pro Leu Leu Ser		Leu Gly Leu Tyr Pro		Cys Trp Leu Leu Phe	
Asn Val Leu Phe		Trp Lys			

(2) INFORMATION FOR SEQ ID NO: 78:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2347 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:
- (A) LIBRARY: HEARNOT01
  - (B) CLONE: 305841

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78 :

```

CCCTCGAGAA GATGGCGGCG ACTCTGGGAC CCCTTGGGTC GTGGCAGCAG TGGCGGCGAT 60
GTTTGTGCGC TCGGGATGGG TCCAGGATGT TACTCCTTCT TCTTTTGTG GGGTCTGGGC 120
AGGGGCCACA GCAAGTCGGG GCGGGTCAAA CGTTCGAGTA CTTGAAACGG GAGCACTCGC 180
TGTCGAAGCC CTACCAGGGT GTGGGCACAG GCAGTTCCTC ACTGTGGAAT CTGATGGGCA 240
ATGCCATGGT GATGACCCAG TATATCCGCC TTACCCCAGA TATGCAAAGT AAACAGGGTG 300
CCTTGTGGAA CCGGGTGCCA TGTTTCCTGA GAGACTGGGA GTTGCAAGTG CACTTCAAAA 360
TCCATGGACA AGGAAAGAAG AATCTGCATG GGGATGGCTT GGCAATCTGG TACACAAAGG 420
ATCGGATGCA GCCAGGGCCT GTGTTTGGAA ACATGGACAA ATTTGTGGGG CTGGGAGTAT 480
TTGTAGACAC CTACCCCAAT GAGGAGAAGC AGCAAGAGCG GGTATTCCCC TACATCTCAG 540
CCATGGTGAA CAACGGCTCC CTCAGCTATG ATCATGAGCG GGATGGGCGG CCTACAGAGC 600
TGGGAGGCTG CACAGCCATT GTCCGCAATC TTCATTACGA CACCTTCCTG GTGATTGCT 660
ACGTCAAGAG GCATTTGACG ATAATGATGG ATATTGATGG CAAGCATGAG TGGAGGGACT 720
GCATTGAAGT GCCCGGAGTC CGCCTGCCCC GCGGCTACTA CTTGCGCACC TCCTCCATCA 780
CTGGGGATCT CTCAGATAAT CATGATGTCA TTTCTTGAA GTTGTTTGA CTGACAGTGG 840

```

PF-0459 US

AGAGAACCCC AGAAGAGGAA AAGCTCCATC GAGATGTGTT CTTGCCCTCA GTGGACAATA 900  
TGAAGCTGCC TGAGATGACA GCTCCACTGC CGCCCCTGAG TGGCCTGGCC CTCTTCCTCA 960  
TCGTCTTTTT CTCCCTGGTG TTTTCTGTAT TTGCCATAGT CATTGGTATC ATACTCTACA 1020  
ACAAATGGCA GGAACAGAGC CGAAAGCGCT TCTACTGAGC CCTCCTGCTG CCACCACTTT 1080  
TGTGACTGTC ACCCATGAGG TATGGAAGGA GCAGGCACTG GCCTGAGCAT GCAGCCTGGA 1140  
GAGTGTTCCTT GTCTCTAGCA GCTGGTTGGG GACTATATTC TGTCCTGGA GTTTTGAATG 1200  
CAGGGACCCC GCATTCCCAT GGTGTGTCAT GGGGACATCT AACTCTGGTC TGGGAAGCCA 1260  
CCCACCCAG GGCAATGCTG CTGTGATGTG CCTTTCCTG CAGTCCTTCC ATGTGGGAGC 1320  
AGAGGTGTGA AGAGAATTTA CGTGGTTGTG ATGCCAAAAT CACAGAACAG AATTTCATAG 1380  
CCCAGGCTGC CGTGTGTTTT GACTCAGAAG GCCCTTCTAC TTCAGTTTTG AATCCACAAA 1440  
GAATTAAAAA CTGGTAACAC CACAGGCTTT CTGACCATCC ATTCGTTGGG TTTTGCATTT 1500  
GACCCAACCC TCTGCCTACC TGAGGAGCTT TCTTTGAAA CCAGGATGGA AACTTCTTCC 1560  
CTGCCTTACC TTCCTTTCAC TCCATTCAAT GTCCTCTCTG TGTGCAACCT GAGCTGGGAA 1620  
AGGCATTTGG ATGCCTCTCT GTTGGGGCCT GGGGCTGCAG AACACACCTG CGTTTCACTG 1680  
GCCTTCATTA GGTGGCCCTA GGGAGATGGC TTTCTGCTTT GGATCACTGT TCCCTAGCAT 1740  
GGGTCTTGGG TCTATTGGCA TGTCCATGGC CTTCCCAATC AAGTCTCTTC AGGCCCTCAG 1800  
TGAAGTTTGG CTAAAGGTTG GTGTAAAAAT CAAGAGAAGC CTGGAAGACA TCATGGATGC 1860  
CATGGATTAG CTGTGCAACT GACCAGCTCC AGGTTTGATC AAACCAAAAG CAACATTTGT 1920  
CATGTGGTCT GACCATGTGG AGATGTTTCT GGACTTGCTA GAGCCTGCTT AGCTGCATGT 1980  
TTTGTAGTTA CGATTTTTGG AATCCCACTT TGAGTGCTGA AAGTGTAAAG AAGCTTTCTT 2040  
CTTACACCTT GGGCTTGGAT ATTGCCCAGA GAAGAAATTT GGCTTTTTTT TTCTTAATGG 2100  
ACAAGAGACA GTTGCTGTTT TCATGTTCCA AGTCTGAGAG CAACAGACCC TCATCATCTG 2160  
TGCCTGGAAG AGTTCCTGT CATTGAGCAG CACAGCCTGA GTGCTGGCCT CTGTCAACCC 2220  
TTATTCCACT GCCTTATTTG ACAAGGGGTT ACATGCTGCT CACCTTACTG CCCTGGGATT 2280  
AAATCAGTTA CAGGCCAGAG TCTCCTTGGA GGGCCTGGAA CTCTGAGTCC TCCTATGAAC 2340  
CTCTGTA 2347

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1529 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: EOSIHET02  
(B) CLONE: 322866

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79 :

CCCACGCGTC CGCCAGCCTT GTCTCGGCCA CCTCAAGGAT AATCACTAAA TTCTGCCGAA 60  
AGGACTGAGG AACGGTGCCT GGAAAAGGGC AAGAATATCA CGGCATGGGC ATGAGTAGCT 120  
TGAAACTGCT GAAGTATGTC CTGTTTTTCT TCAACTTGCT CTTTGGATC TGTGGCTGCT 180  
GCATTTTGGG CTTTGGGATC TACCTGCTGA TCCACAACAA CTTGCGAGTG CTCTCCATA 240  
ACCTGCCCTC CCTCACGCTG GGCAATGTGT TTGTCATCGT GGGCTCTATT ATCATGGTAG 300  
TTGCCTTCCT GGGCTGCATG GGCTCTATCA AGGAAAACAA GTGTCTGCTT ATGTCGTTCT 360  
TCATCCTGCT GCTGATTATC CTCCTTGCTG AGGTGACCTT GGCCATCCTG CTCTTTGTAT 420  
ATGAACAGAA GCTGAATGAG TATGTGGCTA AGGGTCTGAC CGACAGCATC CACCGTTACC 480  
ACTCAGACAA TAGCACCAAG GCAGCGTGGG ACTCCATCCA GTCATTTCTG CAGTGTTGTG 540  
GTATAAATGG CACGAGTGAT TTGGACAGTG GCTCACCAGC ATCTTGCCCC TCAGATCGAA 600  
AAGTGGAGGG GTGCTATGCG AAAGAAGACT TTGGTTTCAT TCAATTCCTT GTATATCGGA 660  
ATCATCACCA TCTGTGTATG TGTGATTGAG GTGTTGGGGG ATGTCCTTTG CACTGACCCT 720  
GAACTGCCAG ATTGACAAAA CCAGCCAGAC CATAGGGCTA TGATCTGCAG TAGTTCTGTG 780  
GTGAAGAGAC TTGTTTCATC TCCGGAAATG CAAAACCATT TATAGCATGA AGCCCTACAT 840  
GATCACTGCA GGATGATCCT CCTCCCATCC TTTCCCTTTT TAGGTCCCTG TCTTATACAA 900  
CCAGAGAAGT GGGTGTGGC CAGGCACATC CCATCTCAGG CAGCAAGACA ATCTTTCCT 960  
CACTGACGGC AGCAGCCATG TCTCTCAAAG TGGTGAACT AATATCTGAG CATCTTTTAG 1020  
ACAAGAGAGG CAAAGACAAA CTGGATTTAA TGGCCCAACA TCAAAGGGTG AACCAGGAT 1080  
ATGAATTTTT GCATCTTCCC ATTGTGCAAT TAGTCTCCAG CCTCTAAATA ATGCCAGTC 1140  
TTCTCCCCAA AGTCAAGCAA GAGACTAGTT GAAGGGAGTT CTGGGGCCAG GCTCACTGGA 1200  
CCATTGTCAC AACCCTCTGT TTCTCTTTGA CTAAGTGCCC TGGCTACAGG AATTACACAG 1260  
TTCTCTTTCT CCAAAGGGCA AGATCTCATT TCAATTTCTT TATTAGAGGG CCTTATTGAT 1320  
GTGTTCTAAG TCTTTCCAGA AAAAACTAT CCAGTGATTT ATATCCTGAT TTCAACCACT 1380  
CACTTAGCTG ATAATCACAG TAAGAAGACT TCTGGTATTA TCTCTCTATC AGATAAGATT 1440  
TTGTTAATGT ACTATTTTAC TCTTCAATAA ATAAAACAGT TTATTATCTC AAAAAAAAAA 1500  
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1529

## (2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4387 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: BEPINOT01  
 (B) CLONE: 546656

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80 :

GCATCCCCGC TTCCGGGTTA GGCCGTTCTT GCCCCCCCC TCCTCTCCTC CCTTCGGACC 60  
 CATAGATCTC AGGCTCGGCT CCCC GCCCGC CGCAGCCCAC TGTTGACCCG GCCCGTACTG 120  
 CGGCCCCGTG GCCACCATGT CCCTGCACGG CAAACGGAAG GAGATCTACA AGTATGAAGC 180  
 GCCCTGGACA GTCTACGCGA TGAAGTGGAG TGTGCGGCCC GATAAGCGCT TTCGCTTGGC 240  
 GCTGGGCAGC TTCGTGGAGG AGTACAACAA CAAGGTTTCT CTTGTTGGTT TAGATGAGGA 300  
 GAGTTCAGAG TTTATTTGCA GAAACACCTT TGACCACCCA TACCCACCA CAAAGCTCAT 360  
 GTGGATCCCT GACACAAAAG GCGTCTATCC AGACCTACTG GCAACAAGCG GTGACTATCT 420  
 CCGTGTGTGG AGGGTTGGTG AAACAGAGAC CAGGCTGGAG TGTTTGCTAA ACAATAATAA 480  
 GAACTCTGAT TTCTGTGCTC CCCTGACCTC CTTTGACTGG AATGAGGTGG ATCCTTATCT 540  
 TTTAGGTACC TCAAGCATTG ATACGACATG CACCATCTGG GGGCTGGAGA CAGGGCAGGT 600  
 GTTAGGGCGA GTGAATCTCG TGTCTGGCCA CGTGAAGACC CAGCTGATCG CCCATGACAA 660  
 AGAGGTCTAT GATATTGCAT TTAGCCGGGC CGGGGGTGGC AGGGACATGT TTGCCTCTGT 720  
 GGGTGCTGAT GGCTCGGTGC GGATGTTTGA CCTCCGCCAT CTAGAACACA GCACCATCAT 780  
 TTACGAAGAC CCACAGCATC ACCCACTGCT TCGCCTCTGC TGGAACAAGC AGGACCCTAA 840  
 CTACCTGGCC ACCATGGCCA TGGATGGAAT GGAGGTGGTG ATTCTAGATG TCCGGGTTCC 900  
 CTGCACACCT GTCGCCAGGT TAAACAACCA TCGAGCATGT GTCAATGGCA TTGCTTGGGC 960  
 CCCACATTCA TCCTGCCACA TCTGCACTGC AGCGGATGAC CACCAGGCTC TCATCTGGGA 1020  
 CATCCAGCAA ATGCCCCGAG CCATTGAGGA CCCTATCCTG GCCTACACAG CTGAAGGAGA 1080  
 GATCAACAAT GTGCAGTGGG CATCAACTCA GCCCGACTGG ATCGCCATCT GCTACAACAA 1140  
 CTGCCTGGAG ATACTCAGAG TGTAAGTTTG GTGGCGCTGT GCCCACGAGG CAGGGGCTTT 1200  
 TGTATTTCTT GCCTCTGCCC CACCCCCAAA GTAAGAAGAA ACATGTTTCC AGTGGCCAGT 1260  
 ATGTCTTTCA TTGCTTTGCA CCCACTGTTA CCAGAAGCTG CTCTAGGAGT TCCTGGCCAG 1320



TCACCCCATC GCCCTCTGTG GCAGACTCAG TGCTGTGTGG CGCCTCCTCA GCCCAGGGCT 1380  
 GAGTTTTAAG ATTTTCTCTC CTTTCCTCTT CTCCTTTGGT TCCTCAATTA AAAAATGTGT 1440  
 GTATATTTGT TTGTCAGGCG TTGTGTTGAG GAGCAGTTCA CGCACTGGCT GTGTCTATTC 1500  
 CTCTGCCCAG GTGTCTCTGT TTGCTGCCCCA AGGCAGCAGT TCATGTCTCG TCCATGTCCA 1560  
 TGTTCTGTGT AGCACTTACG TGGGAACAAA TACCAATTTG TCTTTTCTCC TAGTATCAGT 1620  
 GTGTTAACA AATTTTAACT TTGTATATTT GTTATCTATC AGGCTAATTT TTTTATGAAA 1680  
 AGAATTTTAC TCTCCTGCTT CATTTCTTTG TCTTATAGTC CTCCCTCTTT GCACCTTCTT 1740  
 CTCTTCCCTC AGTGCCCTGGA GCTGGTACTG GGCCCTGGG CCCCATGAGC AGTTTGCCTT 1800  
 CTTGAGTCAC TGCCTGTGTA GTACATACCT GACCGGGAGT CCAAACCACC TTGGTGCTCT 1860  
 GAAGTCCACT GACTCATCAC ACCTTTCTTA GCCTGGCTCC TCTCAAGGGC ATTCTGGGCT 1920  
 TGTAAACAGA CATAGGAAGC CTCTGTTTAC CCTGAAGCAC CACTGTCCAG CCCATTGGTT 1980  
 CCCACTGGCA GCATGGTAGA GCTGAGAGAA ACAGGCTCTC AGGGTACCTG ACTTGAGGGG 2040  
 AATCGTTTCA TGAAGCTGAA CTTCAAGCAT ATTTCCAGTA CATTCTTTCA GAGTCTGTTT 2100  
 TTCCATCCAA ATATAAGCCC CAGGCCATTC CACTTAGTGT CTTTTCAATG ATAGGCAAGA 2160  
 ATGATATCTG AGTTGAACTT CGGTGCTTCT GTTGTGTTGAG TTTACTGTGC CTGGTGGTAT 2220  
 ATTGGGCATT CTTTGGATTG AGTGTTCTGA GGTGAGAGAG TCTTCCCGAG GCATCCTGTC 2280  
 TGTGCTTCCA ACCCTGAACA AGACCTTACA TGAGAGATGG ACTGATGGAC TGCGGCAATC 2340  
 CTGGGCTGTC AAGTGGATAG ATAGTTAAAA AGCATTATAC TGTGGGTAAT GAAAAGGGAG 2400  
 GAAAAAAAAA GAAGGAAAAG GAATTATAGA CCCCAGGGT CAGCCAGTTA AGAGCTCTAC 2460  
 CCACACCTGT CAACCCCTCT CTCCCCAGT TTAGGTTCTG AGCAGTATTG GACTTGTAGC 2520  
 CTGCAGTTGT CTTTTGACTT GCAGGCCGCA GGTGTCTTTC TGTTATGTGA ATGAGTTCCA 2580  
 TGGAGGGGCA TATGTGTGAT TCCACCGTTA GATGAGCCCT TGGGGCAGGC AGTTTGGGAT 2640  
 GTGCTCTTGG GGGAAAGTTG GCTGTTTCCT TGCCTCTGCT TCCTACCCGA AGGTTTTTAA 2700  
 GTCCCTCTGA ATTGCTCATC TGAGATTAGT AGAGTAGCAG GCCTGAAGGA TGATGGTTTT 2760  
 GTCCTCTTTG GTTCTCACCT GCTTGAGAAG TAAACAGTA ACTTTGTTCT TCTGGGCCCT 2820  
 TAAGCTTTTT TGGTTAAGTC TTCCTTTTCA GAAGTAGATG TCATTATATG CCAAAGTCT 2880  
 AGCTCTTTC TTTACCATAC AGGGACCTGT CCCAAAGAAA AAGGCTCTTT TTTTAGCCAG 2940  
 CATATTTCCC CTTCTACCTT TTTACTTTGT TGTTCTGATT TTAGGACTCT GGCTGGCCAT 3000  
 GTGCTTGTGG TTGCTCTCC TGCAATTGCC ACTGGATTG CACTGCATCG TTTGGAGATA 3060  
 CAAAGCGAGC AGTTCTTGGT CAGAACCCTC CTCTGCTTTT CATTGTGTTT GATAATGGTT 3120  
 ACTGGGTCCT TCTCTCAAGG GTAGCAAGGC CAAGCTGATG GCTGCTTGTT TAGGAGGCCA 3180

PF-0459 US

TCAGTTCCTT CCTGTGGAGA AGGGTCTGAA ATGGAAGTCA GTGGTAGAAG GGGCTGGTCT 3240  
GCTGGGCAGG GCTTACATCC ACTGAGTTCT AAGATTCCTT TCCTGATCTG CACCTACGCC 3300  
TGGTCTGTAT GGTGGAATTT GTCAGCTGGA ACTCAGAAAC AACAACTGA AAAAAAATA 3360  
ATAATTAGAA CATATTTGCA TAAGATAGCT ATTTACTCTG GAAACCAACA ACTTTTGAGA 3420  
TTTCCCTTGC CCTGTGGACG CCCAGCTCCT GTCATCCTTC CTTAGGTCCT GCAGTACAGT 3480  
CTTCCCTGA ATGCCACCGG GGACCCAGGG GGACTCCACC CCCCTAAGCA AGCACACACA 3540  
TACTCACAGT TGATGAGTTG CTGGTCTTTG AGTCCCAGCT CTCTTACCCT CCCTTTACTC 3600  
CACCAGCCCG ACGACCCATG ACTGAGGAGG GGATTTCTAC AGTCTCAGGA TTTAGAAAGT 3660  
CTGTAAGCCA TCCATGCTCC AGAAAGCACC GATCTGTTGT AGTTGCAAAA ACAACTCTGT 3720  
AATTTGTTGA GGTTCCTCAA CTGACAGCCA GCGAGACTGG GTGGGAGGCC CTGGATCTGT 3780  
TCTCCCTGAC TGCGGGAGGA GCAGCCACTA GGACTTTAGC AGGAAGCCCA CATGGAGGCT 3840  
CCGCCAGGCT GTGGCCCAGC TGGTGATGGC CCTTTTGCTC CTGGCAGCCT GAGGCACAGC 3900  
TGCCGTGTATT GTCCTCATCT GTTCTGACTG AAGGATGGAG GTGCTGAATA AATTAGGCCT 3960  
CAGGCCTCTA CCACCAGAGA GCTGGAGAAT GGGTCCACGT CATTCAAGGA CCTGAATTTT 4020  
TTATGCTCAG GAGCATTGGA ATCCTCTTCT TCCAGGGAGG AATTAGCCTG CAAGGTTAGG 4080  
ACTTGAAGAG GGAAGGTATT TAATAACTGG GCGAGGATGG GTGTGGTGGC TCACACCTGT 4140  
AATCCCAGCA TTTTGGGAGG CTGAGGTGGC CAGATCCCAA GGTCAGAAGA TCGAGACCAT 4200  
CCTGGCTAAC ATGGTGAAAC CCCATCTCTA CTAAAAATAC AAAAAAAAT TAGCCGGGGG 4260  
TGGTGGCGGG TACCTGTAGT CCTAGCTACT TGGGAGGCTG AGGCAGGAGA ATGGCGTGAA 4320  
CCTGGGAGGT GGAGCTTGCA GTGAGCCAAG ATCGTCCACT CACTGCAGCC TGGCGACAGA 4380  
GCAAGCG 4387

(2) INFORMATION FOR SEQ ID NO: 81:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2117 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: SYNORAT03  
(B) CLONE: 693453

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81 :

GCCTGAGCGG GAAGCATTGG CGTCCGAGCG ACTTCTAGGA GCCTGGGGTT CGGCGCTATG 60

PF-0459 US

GAGGAGCTCG ATGGCGAGCC AACAGTCACT TTGATTCCAG GCGTGAATTC CAAGAAGAAC 120  
CAAATGTATT TTGACTGGGG TCCAGGGGAG ATGCTGGTAT GTGAAACCTC CTTCAACAAA 180  
AAAGAAAAAT CAGAGATGGT GCCAAGTTGC CCCTTTATCT ATATCATCCG TAAGGATGTA 240  
GATGTTTACT CTCAAATCTT GAGAAAACTC TTCAATGAAT CCCATGGAAT CTTTCTGGGC 300  
CTCCAGAGAA TTGACGAAGA GTTGAAGTGA AAATCCAGAA AATCTCAATT GGTTCGAGTG 360  
AGTAAAAACT ACCGATCAGT CATCAGAGCA TGTATGGAGG AAATGCACCA GGTTGCAATT 420  
GCTGCTAAAG ATCCAGCCAA TGGCCGCCAG TTCAGCAGCC AGGTCTCCAT TTTGTCAGCA 480  
ATGGAGCTCA TCTGGAACCT GTGTGAGATT CTTTTTATTG AAGTGGCCCC AGCTGGCCCT 540  
CTCCTCCTCC ATCTCCTTGA CTGGGTCCGG CTCCATGTGT GCGAGGTGGA CAGTTTGTCTG 600  
GCAGATGTTT TGGGCAGTGA GAATCCAAGC AAACATGACA GCTTCTGGAA CTTGGTGACC 660  
ATCTTGGTGC TGCAGGGCCG GCTGGATGAG GCCCACAGA TGCTCTCAA GGAAGCCGAT 720  
GCCAGCCCCG CCTCTGCAGG CATATGCCGA ATCATGGGGG ACCTGATGAG GACAATGCCC 780  
ATTCTTAGTC CTGGGAACAC CCAGACACTG ACAGAGCTGG AGCTGAAGTG GCAGCACTGG 840  
CACGAGGAAT GTGAGCGGTA CCTCCAGGAC AGCACATTCG CCACCAGCCC TCACCTGGAG 900  
TCTCTCTTGA AGATTATGCT GGGAGACGAA GCTGCCTTGT TAGAGCAGAA GGAAGTTCTG 960  
AGTAATTGGT ATCATTTCTT AGTGAAGTCG CTCTGTACT CCAATCCCAC AGTAAAACCC 1020  
ATTGATCTGC ACTACTATGC CCAGTCCAGC CTGGACCTGT TTCTGGGAGG TGAGAGCAGC 1080  
CCAGAACCCC TGGACAACAT CTTGTTGGCA GCCTTTGAGT TTGACATCCA TCAAGTAATC 1140  
AAAGAGTGCA GCATCGCCCT GAGCAACTGG TGGTTTGTGG CCCACCTGAC AGACCTGCTG 1200  
GACCACTGCA AGCTCCTCCA GTCACACAAC CTCTATTTCT GTTCCAACAT GAGAGAGTTC 1260  
CTCCTGCTGG AGTACGCCTC GGGACTGTTT GCTCATCCCA GCCTGTGGCA GCTGGGGGTC 1320  
GATTACTTTG ATTACTGCCC CGAGCTGGGC CGAGTCTCCC TGGAGCTGCA CATTGAGCGG 1380  
ATACCTCTGA ACACCGAGCA GAAAGCCCTG AAGTGCTGTC GGATCTGTGA GCAGCGGCAG 1440  
ATGACTGAAC AAGTTCGCAG CATTGTGAAG ATCTTAGCCA TGAAAGCCGT CCGCAACAAT 1500  
CGCCTGGGTT CTGCCCTCTC TTGGAGCATC CGTGCTAAGG ATGCCGCCTT TGCCACGCTC 1560  
GTGTCAGACA GGTTCCCTCAG GGATTACTGT GAGCGAGGCT GCTTTTCTGA TTTGGATCTC 1620  
ATTGACAACC TGGGGCCAGC CATGATGCTC AGTGACCGAC TGACATTCCT GGGAAAGTAT 1680  
CGCGAGTTCC ACCGTATGTA CGGGGAGAAG CGTTTTGCCG ACGCAGCTTC TCTCCTTCTG 1740  
TCCTTGATGA CGTCTCGGAT TGCCCTCGG TCTTTCTGGA TGAATCTGCT GACAGATGCC 1800  
TTGCCCCTTT TGGAACAGAA ACAGGTGATT TTCTCAGCAG AACAGACTTA TGAGTTGATG 1860  
CGGTGTCTGG AGGACTTGAC GTCAAGAAGA CCTGTGCATG GAGAATCTGA TACCGAGCAG 1920

PF-0459 US

CTCCAGGATG ATGACATAGA GACCACCAAG GTGGAAATGC TGAGACTTTC TCTGGCACGA 1980  
AATCTTGCTC GGGCAATTAT AAGAGAAGGC TCACTGGAAG GTTCCTGAGA ACTGCTTCAA 2040  
TGTGGTATCT TTGTATGGCA ATGTATATAG ATTTTTTAAA AGAATAAATG TTGTTTGCAA 2100  
AAAAAAAAAA AAAAAAA 2117

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 846 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: BRAITUT03  
(B) CLONE: 866885

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82 :

GGCGGGCGGA GTCTGCAGGA TGGCACCGGA CCCCTGGTTC TCCACATACG ATTCTACTTG 60  
TCAAATTGCC CAAGAAATTG CTGAGAAAAT TCAACAACGA AATCAATATG AAAGAAAAGG 120  
TGAAAAGGCA CCAAAGCTTA CCGTGACAAT CAGAGCTTTG TTGCAGAACC TGAAGGAAAA 180  
GATCGCCCTT TTGAAGGACT TATTGCTAAG AGCTGTGTCA ACACATCAGA TAACACAGCT 240  
TGAAGGGGAC CGAAGACAGA ACCTCTTGGA TGATCTTGTA ACTCGAGAGA GACTACTTCT 300  
GGCATCCTTT AAGAATGAGG GTGCCGAACC AGATCTAATC AGGTCCAGCC TGATGAGTGA 360  
AGAGGCTAAG CGAGGAGCAC CCAACCCTTG GCTCTTTGAG GAGCCAGAGG AGACCAGAGG 420  
CTTGGGTTTT GATGAAATCC GGCAACAGCA GCAGAAAATT ATCCAAGAAC AGGATGCAGG 480  
CCTTGATGCC CTTTCCTCTA TCATAAGTCG CCAAAAACAA ATGGGGCAGG AAATTGGGAA 540  
TGAATTGGAT GAACAAAATG AGATAATTGA CGACCTTGCC AACCTAGTGG AGAACACAGA 600  
TGAAAACTT CGCAATGAAA CCAGGCGGGT AAACATGGTG GACAGAAAGT CAGCCTCTTG 660  
TGGGATGATC ATGGTGATTT TACTGCTGCT TGTGGCTATC GTGGTTGTTG CAGTCTGGCC 720  
GACCAACTGA TGGCAGTAAA GAGACCACCA GCAGTGACAC CTGGCAATGA CAGATGCAAG 780  
CCCAACACCC TTTTGGTACG CAAAACCTGC TCTCAATAAA TTCCCCCAA GCTCTGAAAA 840  
AAAAAA 846

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1011 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: LUNGNOT03  
 (B) CLONE: 1242271

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83 :

GAAAGAGATA ACTGGAAGTT CCTTGATTCA GAAAACAGAT TCAGATGAAG AAGTTGCAAT 60  
 GCTGTTGGAC ACAGTCCAGA AAGTATTTCA GAAAATGTTG GAATGTATTG CACGGAGCTT 120  
 CAGGAAGCAG CCGGAAGAAG GCCTGCGGCT GCTTTATTCT GTTCAGAGGC CTCTTCATGA 180  
 GTTCATTACT GCTGTTTCACT CTCGGCACAC AGACACCCCT GTGCACCGGG GTGTACTTTC 240  
 TACTCTGATC GCTGGGCCTG TGTTTGAGAT AAGTCACCAG CTACGGAAGG TTTCTGACGT 300  
 AGAAGAGCTT ACCCCTCCAG AGCATCTTTC TGATCTTCCA CCATTTTCAA GGTGTTTAAT 360  
 AGGAATAATA ATAAAGTCTT CGAATGTGGT CAGGTCATTT TTGGATGAAT TAAAGGCATG 420  
 TGTGGCTTCT AATGATATTG AAGGCATTGT GTGCCTCACG GCTGCTGTGC ATATTATCCT 480  
 GGTTATTAAT GCAGGTAAAC ATAAAAGCTC AAAAGTGAGG GAGGTTGCAG CCACTGTTCA 540  
 CAGAAACTA AAGACATTCA TGGAAATTAC TTTGGAAGAG GATAGCATTG AAAGATTTCT 600  
 CTATGAATCA TCATCAAGAA CTCTGGGAGA ACTTTTGAAT TCATAACCAA GCCAACATCT 660  
 CCAGACATGT AAAAATAGGG AAAAGTGATT CAAATTGAAA TGCCTGTGTA TTTTCCTATT 720  
 GTTTTTAATG TTAATAACCC ATATAATAGG GAAAGGGTGG GATTTTTTTT TGGGAATGTG 780  
 GGAAGGTGGG GGTTATGGAG GAGATAACTC AAAACTTCTT CAATTTTGCC TAGTGCCTGC 840  
 GTAAATAATA TATTTAATAT AAAGGACTCC AGGTATGAAT GGTGTAGAAA TCCATGATTC 900  
 CAAGAAAAA CACTTTTCTA GCAAACCTGG TTGTTTTTAA AATGACTTTT ATATATGTAA 960  
 TATTGCTTGG AAACATATGAG TAATAAAGCA ATGACAACAT CAAAAAAAAA A 1011

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2478 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: LUNGFET03  
 (B) CLONE: 1255027

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84 :

CCCACGCGTC CGCCACGCG TCCGAGCGC TGTGTTTGCG AGCGGGAGCG AGGGGCGCCG 60  
 GCTGGGGTGT GTGCTCCTGA GCTCTTCAGA AACCAGGCTG CTTTCAGGAA CATTGCTGTG 120  
 GATTCCCAGC TTTCAGACAA CACATGACTA AGACAGATGA GACCACTCTA GTTGCCTCAT 180  
 GGGAAACTCG GGAAAAGACT GCAAAAACAA CATTGTTTCT CCCTTTGGAA TTCTGGAGTT 240  
 ATAAGGCAGA GGTCCCCCAT CTTCCCGAAC TGGCCTATTC CGCTAGAAGC AAGATGGCTG 300  
 AACTCAATAC TCATGTGAAT GTCAAGGAAA AGATCTATGC AGTTAGATCA GTTGTTCCCA 360  
 ACAAAGCAA TAATGAAATA GTCCTGGTGC TCCAACAGTT TGATTTTAAT GTGGATAAAG 420  
 CCGTGCAAGC CTTTGTGGAT GGCAGTGCAA TTCAAGTTCT AAAAGAATGG AATATGACAG 480  
 GCAAAAAGAA GAACAATAAA AGAAAAAGAA GCAAGTCCAA GCAGCATCAA GGCAACAAAG 540  
 ATGCTAAAGA CAAGGTGGAG AGGCCGTGAGG CAGGGCCCCCT GCAGCCGCAG CCACCACAGA 600  
 TTCAAAACGG CCCCATGAAT GGCTGCGAGA AGGACAGCTC GTCCACAGAT TCTGCTAACG 660  
 AAAAACCAGC CCTTATCCCT CGTGAGAAAA AGATCTCGAT ACTTGAGGAA CCTTCAAAGG 720  
 CACTTCGTGG GGTACAGAA GGCAACAGAC TACTGCAACA GAAACTATCC TTAGATGGGA 780  
 ACCCCAAACC TATACATGGA ACAACAGAGA GGTCAGATGG CCTACAGTGG TCAGCTGAGC 840  
 AGCCTTGTA CCAAGCAAG CCTAAGGCAA AACATCTCC TGTTAAGTCC AATACCCCTG 900  
 CAGCTCATCT TGAAATAAAG CCAGATGAGT TGGCAAAGAA AAGAGGCCCA AATATTGAGA 960  
 AATCAGTGAA GGATTTGCAA CGCTGCACCG TTTCTCTAAC TAGATATCGC GTCATGATTA 1020  
 AGGAAGAAGT GGATAGTTCC GTGAAGAAGA TCAAAGCTGC CTTTGCTGAA TTACACAACT 1080  
 GCATCATTGA CAAAGAAGTT TCATTAATGG CAGAAATGGA TAAAGTTAAA GAAGAAGCCA 1140  
 TGGAAATCCT GACTGCTCGT CAGAAGAAAG CAGAAGAACT AAAGAGACTC ACTGACCTTG 1200  
 CCAGTCAGAT GGCAGAGATG CAGCTGGCCG AACTCAGGGC AGAAATTAAG CACTTTGTCA 1260  
 GCGAGCGTAA ATATGACGAG GAGCTCGGGA AAGCTGCCCC GTTTTCCTGT GACATCGAAC 1320  
 AGCTGAAGGC CCAAATCATG CTCTGCGGAG AAATTACACA TCCAAAGAAC AACTATTCCT 1380  
 CAAGAACTCC CTGCAGCTCC CTGCTGCCTC TGCTGAATGC GCACGCAGCA ACCTCTGGGA 1440  
 AACAGAGTAA CTTTTCCCGA AAATCATCCA CTCACAATAA GCCCTCTGAA GGCAAAGCGG 1500  
 CAAACCCCAA AATGGTGAGC AGTCTCCCCA GCACCGCCGA CCCCTCTCAC CAGACCATGC 1560  
 CGGCCAACAA GCAGAATGGA TCTTCTAACC AAAGACGGAG ATTTAATCCA CAGTATCATA 1620  
 ACAACAGGCT AAATGGGCCT GCCAAGTCGC AGGGCAGTGG GAATGAAGCC GAGCCACTGG 1680  
 GAAAGGGCAA CAGCCGCCAC GAACACAGAA GACAGCCGCA CAACGGCTTC CGGCCCAAAA 1740  
 ACAAAGGCGG TGCCAAAAAT CAAGAGGCTT CCTTGGGGAT GAAGACCCCC GAGGCCCCCG 1800

PF-0459 US

CCCATTCTGA AAAGCCCCGG CGAAGGCAGC ACGCTGCAGA CACCTCGGAG GCCAGGCCCT 1860  
TCCGGGGTAG TGTCGGTAGG GTTTCACAGT GCAATCTCTG CCCCACGAGA ATAGAAGTTT 1920  
CCACAGATGC AGCAGTTCTC TCAGTCCCGG CTGTGACGTT GGTGGCCTGA GCTAGGAGGA 1980  
AAAAGAGCAG TTTTCACTCA GTTTTGGTTC CCTGCCCAGG GTGCTGACCC AATTCGCTGC 2040  
CAAAAGAGTG TCAATCAGAA TATACAAATC CCGTATGGTT GTGTCATCCT CTCTTAATCA 2100  
TTTTTACTAA TTCTAATAAT CAGCTCTAGC TTGCTTCATA ATTTTCATGG CTTTGCTTGA 2160  
TCTGTTGATG CTTTCTCTCA TCAAGACTTT GCAGCATTTT AGCCAGGCAG TATTTACTCA 2220  
TTATTAGGAA AATCAAGATG TGGCTGAAGA TCAGAGGCTC AGTTAGCAAC CTGTGTTGTA 2280  
GCAGTGATGT CAGTCCATTG ATTGTCTTTA GAGAGTTAAT GTTACAAAAA AGAATTCTTA 2340  
ATAATCAGAC AAACATGATC TGCTGAGGAC ACATGCGCTT TTGTAGAATT TAACATCTGG 2400  
TGTTTTTCTG AAAAAATATA TATACATATA TTGCTTTATT TGAAACAAAT TAAAATATGC 2460  
TGCATTTGAA AAAAAAAA 2478

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1897 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: TESTTUT02
- (B) CLONE: 1273453

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85 :

TGCACATCTA GCACAAATTG AAGATGATAG AGCTGCGATG GTTATTTCTT GGCATCTGGC 60  
AAGTGACATG GACTGTGTAG TCACCCTAAC CACTGACGCT GCACGTCGTA TCTATGATGA 120  
AACCCAAGGT CGTCAGCAGG TGTTGCCCTT TGATTCTATT TACAAGAAGA CTCTTCCAGA 180  
TTGGAAAAGA TCTCTACCTC ATTTCCGAAA TGGAAAATTG TATTTTAAAC CCATTGGAGA 240  
TCCAGTCTTT GCTCGAGACT TGTTAACATT TCCAGATAAT GTAGAACATT GTGAAACAGT 300  
ATTTGGTATG CTGTTAGGAG ACACCATTAT TTTGGATAAT CTGGATGCGG CCAATCATT 360  
TAGAAAAGAG GTTGTTAAAA TTACACACTG TCCTACACTG CTGACCAGAG ATGGAGATCG 420  
AATTCGAAGT AATGGAAAGT TTGGGGGCCT TCAGAATAAA GCTCCTCCAA TGGATAAACT 480  
TCGGGGAATG GTATTTGGAG CTCCAGTTCC AAAACAGTGT CTGATCTTAG GGGAACAAAT 540  
AGATCTTCTT CAGCAGTATC GTTCTGCTGT GTGCAAATA GACAGTGTGA ATAAGGATCT 600

PF-0459 US

TAACAGTCAA TTAGAGTACC TTCGCACTCC GGATATGAGG AAGAAAAAGC AAGAACTTGA 660  
TGAACATGAG AAAAATCTCA AACTAATAGA GGAAAAACTA GGTATGACTC CCATACGTAA 720  
GTGTAATGAC TCATTGCGTC ATTACCAAAA GGTGAGACG ACAGATTGTC CAGTTCCTCC 780  
TAAAAGAATG AGACGAGAAG CTACAAGACA AAATAGGATT ATAACCAAAA CAGATGTATG 840  
AGAGGTGACA GAGAGAAGAG GCCATTGGTC TCAGTAAGAA TGCCCTGCTT TCTGCATCTC 900  
TGTTTCAGAA GACCAAGAGG GTGACTTACC AGACTGAGTA TTTCTGGGGA CAATACAAGT 960  
ACCTGGGCAT GAATTTCCAT TTCGATTGAG ATGGGACTGG AAACAACCAT TCAATTTTAT 1020  
GAATCTTACT GGACATTATG GATTTACTGG AATTATTCCA GACATTATGC CCTTTGGTTG 1080  
TCACTACCTT GCAAATGTGT AAGAGGAAAA TGTGCTAATG TGGCAGTGAC TGTAAAACTG 1140  
GCACATGGCA TTTATTAATC CTGAAGAAAA GTACATGTAC TATTTTTCAG TATAAATATA 1200  
ATGAACATGT CAGAACTATT TCTTGAAAAC CTTTTTATTA CTTTTCGCTG AATTTATTTA 1260  
ACAAAGATGT TTTGTCTTTT GTGTAAGGGA GGTTCAGAG GCTAGATGTT TAATTGTAAA 1320  
TATGTGAGGA AACTCAATGC AGAATTCAGG ATAAAAATTT TAAAAGCACA GGTATTTGGG 1380  
AATTGAAATG TTAAGATACC CAGAACAACA TTAAATCAAT GAGTGAAGTT GTGACAGTGG 1440  
TAGCATTTCA AATTTCAAAA GACTTATCCT GTGTGTGTGT GTGTGTGTGT ATATATATAT 1500  
ATATATATAT AAATATATAT ATATAAATA TTCAGCAGCA CCAAGTTTTA TAACTATTGT 1560  
TTGTTTGAAT TTATTAATAC TAGAATATGT AGTCTCAGCC TTAATTTTAC ATTTACATTA 1620  
TTTTGTAAAT TTTTATTACT ATTTTAAAGG GGTAAAGAG AACATACATT CTCACATTAG 1680  
TGTAATTTCT GGTAGAAAGT TGCTGCAAAA ACATTTGAAA TGTATATTAA CCTAATGTAT 1740  
GTCATATATA TGTCTTTGTG TAAGTTCAAG ACTATTGATC TGTGAAGTTA TTTTGTAAAG 1800  
ACATACATTT GGTAAGTAAG TTTGTGTCCC AGGAAATGTA TGTGTTTTTA AACCCCTTCT 1860  
AAATATGCAG GCCATTAATA AATAAGATTG TGTCTCA 1897

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1488 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: TESTTUT02
- (B) CLONE: 1275261

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86 :



PF-0459 US

CCCACGCGTC CGGGGACATC CTGTTCTGAG TCAAGATTCC TCCTTCTGAA CATGGGACTT 60  
TCCAGAAGGA CCACAGCTCC TCCCGTGCAT CCACTCGGCC TGGGAGGTTT TGGATTTTGG 120  
CTGTCGAGGG AGTTTGCCTG CCTCTCCAGA GAAAGATGGT CATGAGGCCC CTGTGGAGTC 180  
TGCTTCTCTG GGAAGCCCTA CTTCCCATTA CAGTTACTGG TGCCCAAGTG CTGAGCAAAG 240  
TCGGGGGCTC GGTGCTGCTG GTGGCAGCGC GTCCCCCTGG CTTCCAAGTC CGTGAGGCTA 300  
TCTGGCGATC TCTCTGGCCT TCAGAAGAGC TCCTGGCCAC GTTTTTCGA GGCTCCCTGG 360  
AGACTCTGTA CCATTCCCCG TTCCTGGGCC GAGCCCAGCT ACACAGCAAC CTCAGCCTGG 420  
AGCTCGGGCC GCTGGAGTCT GGAGACAGCG GCAACTTCTC CGTGTTGATG GTGGACACAA 480  
GGGGCCAGCC CTGGACCCAG ACCCTCCAGC TCAAGGTGTA CGATGCAGTG CCCAGGCCCG 540  
TGGTACAAGT GTTCATTGCT GTAGAAAGGG ATGCTCAGCC CTCCAAGACC TGCCAGGTTT 600  
TCTTGTCTCG TTGGGGCCCC AACATCAGCG AAATAACCTA TAGCTGGCGA CGGGAGACAA 660  
CCATGGACTT TGGTATGGAA CCACACAGCC TCTTCACAGA CGGACAGGTG CTGAGCATTT 720  
CCCTGGGACC AGGAGACAGA GATGTGGCCT ATTCCTGCAT TGTCTCCAAC CCTGTCAGCT 780  
GGGACTTGGC CACAGTCACG CCCTGGGATA GCTGTCATCA TGAGGCAGCA CCAGGGAAGG 840  
CCTCCTACAA AGATGTGCTG CTGGTGGTGG TGCCTGTCTC GCTGCTCCTG ATGCTGGTTA 900  
CTCTCTTCTC TGCCTGGCAC TGGTGCCCCT GCTCAGGGAA AAAGAAAAAG GATGTCCATG 960  
CTGACAGAGT GGGTCCAGAG ACAGAGAACC CCCTTGTGCA GGATCTGCCA TAAAGGACAA 1020  
TATGAACTGA TGCCTGGACT ATCAGTAACC CCACTGCACA GGCACACGAT GCTCTGGGAC 1080  
ATAACTGGTG CCTGGAAATC ACCATGGTCC TCATATCTCC CATGGGAATC CTGTCCTGCC 1140  
TCGAAGGAGC AGCCTGGGCA GCCATCACAC CACGAGGACA GGAAGCACCA GCACGTTTCA 1200  
CACCTCCCCC TTCCCTCTCC CATCTTCTCA TATCCTGGCT CTTCTCTGGG CAAGATGAGC 1260  
CAAGCAGAAC ATTCCATCCA GGACACTGGA AGTTCTCCAG GATCCAGATC CATGGGGACA 1320  
TTAATAGTCC AAGGCATTCC CTCCCCACC ACTATTCTA AAGTACTAAC CAACTGGCAC 1380  
CAAGAAAAAA TCCTCACTAA CCGCATCATC CGACAACTAA TAATTCACAC TACATCCAAA 1440  
CATCACTTAG GCGGCGGGGC CGCCGACTGG TTCCGGGCTT AGGGTGGG 1488

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1357 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: COLNNOT16  
 (B) CLONE: 1281682

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87 :

CCGACTTTGT AGCATTTTTA TTTAAGCTAA AACAGAGCAC ATGTATATGT ACATAAGACA 60  
 CATTAAATCT ATAAATACTA TTTATTCATT TTATATAAAC TAATGTAATG GAAAACAAAT 120  
 TCTTATGACT TTGTGGTTTT ATAGATGTTT TAGAAACTTT GTATGTAGGT ATCTACAAAA 180  
 TTAGTTCATT CCCCTGAATA TTTTTCATT CATATTTTGT AGGTCTTGAT GTTTTCAGCC 240  
 TCTGGCGAAT CTTTTTCATT GAATTTGAAC CATTGTGAAA ATCTGTGATG CTGAAGCAGA 300  
 GTGTGTCACA AAGTGATGAG AACATTACTA AAATCCACGG ACGCACTGCG ACCTAAGGGC 360  
 TCAACGGCTG ACTCGGCAGC GGGCAGCCAC CCCACGCTCC CCTGCGGTCA CTCGCACACC 420  
 ACAGCCTGAA GCTCCCCCAG CGCCTGCACC TCGCACACAG CTAAGGTCAA AGTTCAAACG 480  
 CACTCCACAC GGAAGCTCAT TCTATACCCG AAGAGCAGTC TCAGAAAGCA AGATTACTTT 540  
 TGTGTTTTTT AAAAATGAT TCTTTAATGT ATTTTCTAA ACATTCTGAT TGAAGTAGT 600  
 GGATTCCTAA ATGATTCCAA AGTCATCTGT AATTCTTCTG TTTTGTGTTT GTTCTGTCTT 660  
 TTCTTCATTT TGGCTTTGGG TGGGGGGAGG GGCAGGTGAC ACAAAGGATT TTTTTTTTTT 720  
 TTTTTTTTTA ATTTTGGAA TCTTTTCCAA TAACCAGCTA AAGATTTGCA CTGAAATACA 780  
 ACTTGTATGC CTTTTCATT TTTAAAGCCT GCTTCCTGGA TTTAAGCAGA GTGATAGTGT 840  
 TCAAAGAGCC AGTTCAGCCT GTAACATATT TGAAAAAGAT ATGTCTGCAC TTTGAGGTCC 900  
 CTTTTGAATG CCATTCACTA GACCTCTCAA GCATTTTGTT TCATTGCTAC ATCCAAGCGC 960  
 CTCACAAGTC CACAATGCGG GACAGCATCA AAAGCTCAAG ACTTTGAAA AAGCTTGTGG 1020  
 GCTTGCACTG GGGGAGGGAA GGAACAAAA TTTGTGTACT TCTTTGTTTA ATTTAGAAAT 1080  
 AAGGCATCCA AGAGATGCCA TTATTTTCTG TGTTTCAATT GTTGTGCCTT TGAGTTAAAC 1140  
 TGCATTTTTG TCTTTTGGTT GAAATCTGAA ATGTACTGTC CCAATATAAA ACAGTAATTA 1200  
 TTTGACCTTT GCACTGTTTG TCTGGTCCTT TTCAGTTTGA TTGCATATAA ATGTGGAAC 1260  
 TGATAGATCT CTATATTTTT AATGCACTTG TGATAAACTG GCAGCAGGGT TAGACATTAC 1320  
 TTTCAAAGCT TGAGGTAGAC CGAGTCAGCA TGCTAGA 1357

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2330 base pairs  
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: BRSTNOT07  
(B) CLONE: 1298305

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88 :

```

CCTACTTGTT CCCACCTTGG GAGAGGACGA TGACTTGGGA GGGACGCGTG AAGGGAGAAG 60
GGGTCTCTCC ATGAGGCTGA GGATGGCCTG AACCTGGAGC AGCGGACCAG GCAGACGGGC 120
TGAAGTGGGG TCCCAAATTC CATGTCCAGA GGTGTGGGGA GCCTGCCTCC CTAGCTCCTG 180
GCCCCTGCCA GGGGCTTACA TCAAAACACC TCAGAGGGCT GCCCTCCAGA GGCTGCACCC 240
AGAACAGTGG GACATGAGCA GGGGTGTGGG CTTGGAGGGT GAAGAGGATG TGGTCCTATC 300
AGATGCTGGG CCTCCTCAGC CATAGCCCCC TGCTCCTACC CCCTGACTGG CTCTTGTTGTC 360
CTCACCTCTC ACCCTCTCCT TCCTGGGAGG CCCTGGGAGG TGATCATTGA CACCCAGCCA 420
AGCAGACAGC TCGGGGTGCC CAAGCCCTTG CTGGGCCTGC GCGTGAGGAG TCCCACTGCT 480
TCTAAAGGAA GTCCTGGGCA GGAGGTGGCT TTGGTGGTTG GTTCCAAAGT TGAAAAATGCT 540
TGCAGTTTGA CCTTAGAAGA AGTGGGAAGA AGAAGGAGCT CTACAGGGTC AGCTTTGTTT 600
GATTTGTCCA GTCTAAGAAG TCCCATTGCC AAAGCTTTCT GCAGGAGGGT GAATGCCGCA 660
GCTTGGCAGC CCCTGGGTTT CTCTTGGAAG TGGTCAGTTT CCCCTCAAAG TACCCAAAGT 720
AGCCTTGGCT TGAGTTTTTG TCCTTGCCCTC CTTTTTAGAG AAGAGGGCAT TTAGACTGCA 780
TTTTCTCTGG TAAAGAAGGT TAAAGCAAAT GTTTATTGCC TTTTCTAGTG AACTAACTCG 840
TAGAGATGTT CTCAGCAGGA AGACAGTCTT AGCACTGTCA CTTAGCAGAT TGCACTTAAG 900
TCCCTTGTGC TGGCCAGATG GCGTGGCTGG TTGCCTTAAT ATGTCCCAGG ACCCCTGACA 960
GGGCTGCCTG GCCTCTCCCT CGTGCTCCTC AAGAGCCCAG TCCATACACT GTGGATGTCA 1020
TTGCTGTGCG GTTAGGAAGT CTTGTCCTAG AACGCCCTGG CTGGTATGAC CACAGTTCAT 1080
GGCGGCTCTT CTCGCTTGGG TCATGGTCAT CTTCCAGCAC CTGCTGTGCT GGGGAAGGCCG 1140
AGGATGGGGG CCCAGCACTG TCCAGGCCCTG CTGGGGCCTG GCTGGGAGTC CTGTGGGCAG 1200
CATGGAACAT GCAGCTGGGC TTCCTGTGAC CAGGCACCCT CTGGCACTGT TGCTTGCCCT 1260
GTGCCCTGGA CCTTTTCCTG CCCTTCTCCT TCCTCTGCTC CCTTGGGGCT ACCCCTTGGC 1320
CCCTCCTGGT CTGTGCAAAC TCCCTCAGGG AGCCCCCTG CCCTGTAGCT CTCACTTAAC 1380
TTCCTAGGGG CTGCTGAGCC CACCCAGAGG TTGTTGGAGT TCAGCGGGGC AGCTTGCTCTC 1440
CCTTGTCAGC AGGGGCGTAA GGGCTGGGTT TGGCCATACA AGGTTGGCTA CGCCCTCAAT 1500
CCCTGACCGT TCCAGGCACT GAGCTGGGCA CCCACGGAAG GACATGCTGT CCAGACTGTG 1560

```

PF-0459 US

ATGACTGCCA GCACAGGGCA TCTCGGGCTT GGCTGGTCTG CGAGGCCTTG CCCCTGTGGA 1620  
ACTCTGGGTT CCTGTTTTCT CAGTCTTTTT GCGGCTTTGC TGTGGTTGGC AGCTGCCGTA 1680  
CTCCAGGCTT GTGTCGGCCA CTCAGATGAG GGCTGTGGTG CGAGCCAGTG CAGGAGAGCT 1740  
GCGCTTGGGA TTGTGCCCTC TCCTGTGTCT GTCCTCCGGA CCTACCCAGG TCTCCACCAT 1800  
CAGGACCCTG TCTTTGGGTT TAGAAGACCA AGTATGGGGA AAACCAGACA CCAGCCTCTG 1860  
CAGCAATGGG TCCCTCTAGC CTGTGGACAC CAGCTGGGGG ATCCAGGGTC AGGCCCCCTC 1920  
CTCTCCCCAG TTTCCCTCTG CTGTGGGTTC TGGGCTGTCA TGTCTCCACC ACTTAAGGAT 1980  
GTCTTTACAC TGACTTCAGG ATAGATGCTG GGATGCCTGG GCATGGCCAC ATGTTACATG 2040  
TACAGAACTT TGTCTACAGC ACAAATTAAG TTATATAAAC ACAGTGA CTG 2100  
CTGATCTACT ATAAGGTATT CTATATTTAT ATGACTTCAG AGACGCGTAT GTAATAAAGG 2160  
ACGCCCTCCC TCCAGTGTCC ACATCCAGTT CACCCCAGAG GGTGCGGCAG GTTGACATAT 2220  
TTATTTTTGT CTATTCTGTA GGCTTCCATG TCCAGAATCC TGCTTAAGGT TTTAGGGTAC 2280  
CTTCAGTACT TTTTGCAATA AAAGTATTTT CTATCCAAAA AAAAAAAAAA 2330

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2729 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGNOT12
- (B) CLONE: 1360501

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89 :

CTACACCTTT TCCATTTGCT AATAAGGCCG TGCCAGGCTG GGAGGGAATT GTCCCTGCCT 60  
GCTTCTGGAG AAAGAAGATA TTGACACCAT CTACGGGCAC CATGGAAGTCT CTTCAAGTGA 120  
CCATTCTTTT TCTTCTGCCC AGTATTTGCA GCAGTAACAG CACAGGTGTT TTAGAGGCAG 180  
CTAATAATTC ACTTGTTGTT ACTACAACAA AACCATCTAT AACAACACCA AACACAGAAT 240  
CATTACAGAA AAATGTTGTC ACACCAACAA CTGGAACAAC TCCTAAAGGA ACAATCACCA 300  
ATGAATTACT TAAAATGTCT CTGATGTCAA CAGCTACTTT TTTAACAAGT AAAGATGAAG 360  
GATTGAAAGC CACAACCACT GATGTCAGGA AGAATGACTC CATCATTTCA AACGTAACAG 420  
TAACAAGTGT TACACTTCCA AATGCTGTTT CAACATTACA AAGTTCCAAA CCCAAGACTG 480  
AAACTCAGAG TTCAATTAAA ACAACAGAAA TACCAGGTAG TGTTCTACAA CCAGATGCAT 540

PF-0459 US

CACCTTCTAA AACTGGTACA TTAACCTCAA TACCAGTTAC AATTCCAGAA AACACCTCAC 600  
AGTCTCAAGT AATAGGCACT GAGGGTGGAA AAAATGCAAG CACTTCAGCA ACCAGCCGGT 660  
CTTATTCCAG TATTATTTTG CCGGTGGTTA TTGCTTTGAT TGTAATAACA CTTTCAGTAT 720  
TTGTTCTGGT GGGTTTGTAC CGAATGTGCT GGAAGGCAGA TCCGGGCACA CCAGAAAATG 780  
GAAATGATCA ACCTCAGTCT GATAAAGAGA GCGTGAAGCT TCTTACCGTT AAGACAATTT 840  
CTCATGAGTC TGGTGAGCAC TCTGCACAAG GAAAAACCAA GAACTGACAG CTTGAGGAAT 900  
TCTCTCCACA CCTAGGCAAT AATTACGCTT AATCTTCAGC TTCTATGCAC CAAGCGTGGA 960  
AAAGGAGAAA GTCCTGCAGA ATCAATCCCG ACTTCCATAC CTGCTGCTGG ACTGTACCAG 1020  
ACGTCTGTCC CAGTAAAGTG ATGTCCAGCT GACATGCAAT AATTTGATGG AATCAAAAAG 1080  
AACCCCGGGG CTCTCCTGTT CTCTCACATT TAAAAATTCC ATTACTCCAT TTACAGGAGC 1140  
GTTCTAGGA AAAGGAATTT TAGGAGGAGA ATTTGTGAGC AGTGAATCTG ACAGCCCAGG 1200  
AGGTGGGCTC GCTGATAGGC ATGACTTTCC TTAATGTTTA AAGTTTTCCG GGCCAAGAAT 1260  
TTTTATCCAT GAAGACTTTC CTACTTTTCT CGGTGTTCTT ATATTACCTA CTGTTAGTAT 1320  
TTATTGTTTA CCACTATGTT AATGCAGGGA AAAGTGCAC GTGTATTATT AAATATTAGG 1380  
TAGAAATCAT ACCATGCTAC TTTGTACATA TAAGTATTTT ATTCCTGCTT TCGTGTTACT 1440  
TTTAATAAAT AACTACTGTA CTCAATACTC TAAAAATACT ATAACATGAC TGTGAAAATG 1500  
GCAATGTTAT TGTCTTCCTA TAATTATGAA TATTTTGGGA TGGATTATTA GAATACATGA 1560  
ACTCACTAAT GAAAGGCATT TGTAATAAGT CAGAAAGGGA CATAGGATTC ACATATCAGA 1620  
CTGTTAGGGG GAGAGTAATT TATCAGTTCT TTGGTCTTTC TATTTGTCAT TCATACTATG 1680  
TGATGAAGAT GTAAGTGCAA GGGCATTAT AACACTATAC TGCAATTCATT AAGATAATAG 1740  
GATCATGATT TTTCATTAAC TCATTTGATT GATATTATCT CCATGCATTT TTTATTTCTT 1800  
TTAGAAATGT AATTATTTGT TCTAGCAATC ATTGCTAACC TCTAGTTTGT AGAAAATCAA 1860  
CACTTTATAA ATACATAATT ATGATATTAT TTTTCATTGT ATCACTGTTC TAAAAATACC 1920  
ATATGATTAT AGCTGCCACT CCATCAGGAG CAAATCTTC TGTTAAAAGC TAACTGATCA 1980  
ACCTTGACCA CTTTTTTGAC ATGTGAGATC AAAGTGTCOA GTTGGCTGAG GTTTTTTGGA 2040  
AAGCTTTAGA ACTAATAAGC TGCTGGTGGC AGCTTTGTAA CGTATGATTA TCTAAGCTGA 2100  
TTTTGATGCT AAATTATCTT AGTGATCTAA GGGGCAGTTT AGTGAAGATG GAATCTTGTA 2160  
TTTAAAATAG CCTTTTAAAA TTTGTTTTGT GGTGATGTAT TTTGACAACT TCCATCTTTA 2220  
GGAGTTATAT AATCACCTTG ATTTTAGTTT CCTGATGTTT GGACTATTTA TAATCAAGGA 2280  
CACCAAGCAA GCATAAGCAT ATCTATATTT CTGACTGGTG TCTCTTTGAG AAGGATGGGA 2340  
AGTAGAAAAA AAAAAAGAA AGAAAGGAAA GGAAGAGAGG AGAGAAGAAG GCAGGGATCT 2400

PF-0459 US

CCACTATGTA TGTTTTCACT TTAGAACTGT TGAGCCCATG CTTAATTTTA ATCTAGAAGT 2460  
CTTTAAATGG TGAGACAGTG ACTGGAGCAT GCCAATCAGA GAGCATTTGT CTTCAGAAAA 2520  
AAAAAAATC TGAGTTTGAG ACTAGCCTGG CCAACATGTT GAAACCCCAT ATCTACTAAA 2580  
AATACAAAAA TTAGCCTGGT GTGGTGGCGC ACGCCTGTAG TCCCAGCTAC TCTGGAGCCT 2640  
GAGGAACGTG AATCGCTTGA ACCCAGAAGA CAGAGGTTGC AGTGAGCTGA GATGGCACTA 2700  
TTGCACTCCA GACTGGTGAC ACACGCAGA 2729

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1386 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGNOT12
- (B) CLONE: 1362406

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90 :

GGCCCCTGCA CTGCTCCTGA TCCCTGCTGC CCTCGCCTCT TTCATCCTGG CCTTTGGCAC 60  
CGGAGTGGAG TTCGTGCGCT TTACCTCCCT TCGGCCACTT CTTGGAGGGA TCCCGGAGTC 120  
TGGTGGTCCG GATGCCCGCC AGGGATGGCT GGCTGCCCTG CAGACCGCAG CATCCTTGCC 180  
CCCCTGGCAT GGGATCTGGG GCTCCTGCTT CTATTTGTTG GGCAGCACAG CCTCATGGCA 240  
GCTGAAAGAG TGAAGGCATG GACATCCCGG TACTTTGGGG TCCTTCAGAG GTCACTGTAT 300  
GTGGCCTGCA CTGCCCTGGC CTTGCAGCTG GTGATGCGGT ACTGGGAGCC CATAACCAAA 360  
GGCCCTGTGT TGTGGGAGGC TCGGGCTGAG CCATGGGCCA CCTGGGTGCC GCTCCTCTGC 420  
TTTGTGCTCC ATGTCATCTC CTGGCTCCTC ATCTTTAGCA TCCTTCTCGT CTTTGACTAT 480  
GCTGAGCTCA TGGGCCTCAA ACAGGTATAC TACCATGTGC TGGGGCTGGG CGAGCCTCTG 540  
GCCCTGAAGT CTCCCCGGGC TCTCAGACTC TTCTCCCACC TGCGCCACCC AGTGTGTGTG 600  
GAGCTGCTGA CAGTGCTGTG GGTGGTGCCT ACCCTGGGCA CGGACCGTCT CCTCCTTGCT 660  
TTCCTCCTTA CCCTCTACCT GGGCCTGGCT CACGGGCTTG ATCAGCAAGA CCTCCGCTAC 720  
CTCCGGGCCC AGCTACAAAG AAAACTCCAC CTGCTCTCTC GGCCCCAGGA TGGGGAGGCA 780  
GAGTGAGGAG CTCACTCTGG TTACAAGCCC TGTTCCTCCT CTCCCCTGA ATTCTAAATC 840  
CTTAACATCC AGGCCCTGGC TGCTTCATGC CAGAGGCCCA AATCCATGGA CTGAAGGAGA 900  
TGCCCCCTTCT ACTACTTGAG ACTTTATTCT CTGGGTCCAG CTCCATACCC TAAATTCTGA 960

PF-0459 US

GTTTCAGCCA CTGAACTCCA AGGTCCACTT CTCACCAGCA AGGAAGAGTG GGGTATGGAA 1020  
GTCATCTGTC CCTTCACTGT TTAGAGCATG ACACTCTCCC CCTCAACAGC CTCCTGAGAA 1080  
GGAAAGGATC TGCCCTGACC ACTCCCCTGG CACTGTTACT TGCCTCTGCG CCTCAGGGGT 1140  
CCCCTTCTGC ACCGCTGGCT TCCACTCCAA GAAGGTGGAC CAGGGTCTGC AAGTTCAACG 1200  
GTCATAGCTG TCCCTCCAGG CCCC AACCTT GCCTCACCAC TCCCGGCCCT AGTCTCTGCA 1260  
CCTCCTTAGG CCCTGCCTCT GGGCTCAGAC CCCAACCTAG TCAAGGGGAT TCTCCTGCTC 1320  
TTAACTCGAT GACTTGGGGC TCCCTGCTCT CCCGAGGAAG ATGCTCTGCA GGAAAATAAA 1380  
AGTCAG 1386

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 542 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: LATRTUT02  
(B) CLONE: 1405329

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91 :

CCCGGGCCAT GCAGCCTCGG CCCC GCGGGC GCCCGCCGCG CACCCGAGGA GATGAGGCTC 60  
CGCAATGGCA CCTTCCTGAC GCTGCTGCTC TTCTGCCTGT GCGCCTTCCT CTCGCTGTCC 120  
TGGTACGCGG CACTCAGCGG CCAGAAAGGC GACGTTGTGG ACGTTTACCA GCGGGAGTTC 180  
CTGGCGCTGC GCGATCGGTT GCACGCAGCT GAGCAGGAGA GCCTCAAGCG CTCCAAGGAG 240  
CTCAACCTGG TGCTGGACGA GATCAAGAGG GCCGTGTCAG AAAGGCAGGC GCTGCGAGAC 300  
GGAGACGGCA ATCGCACCTG GGGCCGCTA ACAGAGGACC CCCGATTGAC GCCGTGGAAC 360  
GGCTCACACC GGCACGTGCT GCACCTGCCC ACCGTCTTCC ATCACCTGCC ACACCTGCTG 420  
GCCAAGGAGA GCAGTCTGCA GCCCGCGGTG CGCGTGGGCC AGGGCCGCAC CGGAGTGTCG 480  
GTGGTGATGG GCATCCCGAG CGTGCGGCGC GAGGTGCACT CGTACCTGAC TGACACTCTG 540  
CA 542

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:

PF-0459 US

- (A) LENGTH: 772 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAINOT12
- (B) CLONE: 1415223

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92 :

CGAGCCCGGA GTGCGGACAC CCCCGGGATG CTTGCGCCCC AGAGGACCCG CGCCCCAAGC 60  
CCCCGCGCCG CCCCAGGCC CACCCGAGC ATGCTGCCTG CAGCCATGAA GGCCTCGGC 120  
CTGGCGCTGC TGGCCGTCCT GCTGTGCTCG GCGCCGCTC ATGGCCTGTG GTGCCAGGAC 180  
TGCACCCTGA CCACCAACTC CAGCCATTGC ACCCCAAAGC AGTGCCAGCC GTCCGACACG 240  
GTGTGTGCCA GTGTCCGAAT CACCGATCCC AGCAGCAGCA GGAAGGATCA CTCGGTGAAC 300  
AAGATGTGTG CCTCCTCCTG TGA TCTCGTT AAGCGACACT TTTTCTCAGA CTATCTGATG 360  
GGGTTTATTA ACTCTGGGAT CTAAAGGTC GACGTGGACT GCTGCGAGAA GGATTTGTGC 420  
AATGGGGCGG CAGGGGCGAG GCACAGCCCC TGGGCCCTGG CCGGGGGGCT CCTGCTCAGC 480  
CTGGGGCCTG CCCTCCTCTG GGCTGGGCCC TGATGTCTCC TGCTTCCCAC GGGGCTTCTG 540  
AGCTTGCTCC CCTGAGCCTG TGGCTGCCCT CTCCCAGCC TGGCGTGGCT GGGGCTGGGG 600  
GCAGCCTTGG GCCAGCTCCG TGGCTGTGGC CTGTGGGTCT GAATTCTTCC CCGACGTGAA 660  
GCCTNCCTGT CTCTCCGGCA GCTCTGAGTC CCAGGCAGCT GGACATTCCA GGGGAACAAG 720  
CCATTNGGCA GGAGGGCTGG GATGAGGTTG GGGGGGACCG GAGGTCCCGG AG 772

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1738 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAINOT12
- (B) CLONE: 1416553

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93 :

TGTCCATCCA AAAACCATAA AATCACTGGG TTCCACATCA GCCTCCATGA GGCCAAGCCT 60  
TGTACCTGCA AGCTCTTGGC CTAACCATTC CTCTGTCTC TTCTCTGGCC TGCCTGGGGA 120  
GCCCCGTGAAG GCCGCACGGG TGCCTCCAGC CTGAGACATC AGGGGAGAGC CTGCAGCTGA 180



PF-0459 US

G TTCAGCAGA AAGGAGGAAT CCTGGCCCTC AGGAAGAAGA TAGTCACATG TTTTCTTCC 240  
T TGTCCCCAC AGCCCCCAGA ACAACATTCT CCCTGCTGGC AGCCCTTCCA TGTCTCCAAA 300  
C CTGGGTCAG AGTGAAAGGA CCTTTGGGGG TGGGTGGGAG CAAAGGGCCC ACCTGCTGGT 360  
T GGTGAAAGC AGTGGTGCCG GAGTGCTAGG TACCGCACGA GTAGTGGTGC GGGGGCTTGG 420  
G AAGCAGACC AGGGTTGGAC AAAACCCCAT GAGGGCGGGG AGCTGGAAGA AAAGTCTCTT 480  
G GGGGACCTCT GGGGCAAGGA GCTGAGAAGT CCTGCAGCAC CAGGTGAGAC TTGCTTACAG 540  
T GGATGCCAC TTCTAGGCCT CTGGACCGCA GATGCCCTCC TCCCTCCTGC ACACCTGGCC 600  
T CCTGGGCCT CCAGGTAAAG AGAGAGAGCC AGCCAGCCC TGTTTCCCCT CAGTCCTCCT 660  
T TGCTCCTGC TGCTTCTCCC AACAGCCCAC TGTTAGGAGG TAGTAGACCC CAGCCTCAAG 720  
G CTCTGACCT TCTTCATGTG GGCACAGAGG GTCCTGACAC TCTGGCAGGG CCTGAGCTGG 780  
G GCAGGCCCTC CCTCAGGGCC AGGGGCGATG GCACCCCGGG GACAGGCAGA CCTCCTTCCT 840  
G CCGTCAGCA CCCCCTTCCT TATCACTGTC TGGTCTCCGA GCTTCGGCTG CAGCCTGAGG 900  
T GTGTCTTGG GCTCCTCAGA GCCTGAAGCA AGCTTTTGGG AGCCTGCAGT CCTCCCAGCT 960  
C CAGTGACAG AGCCTCTCTC TCCAGCCTTT CCCCAGGCAG GAGTTGGGGT TGGGGGCCTC 1020  
T GTCCCTCAT CGCTTACCTT GGAAAGGTGG GAAGCTGGCA ATCTGCACCT TGGGGCCTGG 1080  
G CTCCCCCTC TCTGTGCCAG CGGCTTCCCA GCACCTGGGA GGGGCTGCAG CCCCAGCTGG 1140  
A CTCCAGCCT GTCCCTCTTA GCACTCTAGC TGCCCACTCC AGGGCAGGGA CTCGAAACCC 1200  
C CTCCGTCCT GAGCAGCCAC CTCCAGGGCC CTGTTTGGGA CCACTCTCTC AGTCCCCAGG 1260  
T CCTCAGGGC CCCAGAGCGG GAGGGTCTCC TACCTGGAAG TCCCCCTGAG CTCCAGGGCC 1320  
C AGCCCTACC TGCCAGTGCT GGTGTCAGGG CACTCAACAC CGAGTGTGGG GGCCACGCCC 1380  
C TTGCCATGC CCACGGCCTC CTCCTGTAGC CCCTGCCTGC ACCCACGATG CTGCACGGGC 1440  
C CGCCCTGGT GGGGCTCGGC GAGTAATGTG TTTTGTCCCC AGTTAACCAC CATTCTGCGG 1500  
C CTGGTTCTG CAAGGAACCA GGGCTGCCCC ACCGCCCGCC GTCTGCCGCC CTAGGCTTCC 1560  
T GACTCCATT AGTTCCGACA CTTGTGAAAC TCCGAGAAGT GCTGTGGTCT CAGCAATGCA 1620  
C CTGTTTTGT ACATGATTGT GTAATTTAAA GGTATATAAA TACAAATATA TATATATATC 1680  
A GTTGTGATT GTATGACTGT GGATAAAATC CAGAACTGTG TCAACCTGAA AAAAAAAA 1738

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2100 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: KIDNNOT09  
(B) CLONE: 1418517

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94 :

```

GGGAAAGCGG CGAGTAAGAT GGAAGATGAG GAGGTCGCTG AGAGCTGGGA AGAGGCGGCA 60
GACAGCGGGG AAATAGACAG ACGGTTGGAA AAAAAACTGA AGATCACACA AAAAGAGAGC 120
AGGAAATCCA AATCTCCTCC CAAAGTGCCC ATTGTGATTC AGGACGATAG CCTTCCCGCG 180
GGGCCCCCTC CACAGATCCG CATCCTCAAG AGGCCACCA GCAACGGTGT GGTGAGCAGC 240
CCCAACTCCA CCAGCAGGCC CACCCTTCCA GTCAAGTCCC TAGCACAGCG AGAGGCCGAG 300
TACGCCGAGG CCCGGAAGCG GATCCTGGGC AGCGCCAGCC CCGAGGAGGA GCAGGAGAAA 360
CCCATCCTCG ACAGGCCAAC CAGGATCTCC CAACCCGAAG ACAGCAGGCA GCCCAATAAT 420
GTGATCAGAC AGCCTTTGGG TCCTGATGGG TCTCAAGGCT TCAAACAGCG CAGATAAATG 480
CAGGCAAGAA AAGATGCCGC CGTTGCTGCC GTCACCGCCT CCTGGGTCGT CCGCCACGGG 540
TTGCACTGCC GTGGCAGACA GCTGGACTTG AGCAGAGGGA ACGACCTGAC TTACTTGAC 600
TGTGATCCCC CTTGCTCCGC CCACTGTGAC CTTGAACCCC ATGCACTGTG ACCTCCCCCC 660
TTCTCCCCCT TCCCACTGTG ATTGGCACAT CGACAAGGGC TGTCCCAAGT CAATGGAAAG 720
GGAAAGGGTG GGGGTTAGGG GAAGGTTGGG GGGACCCAGC AAGGACTCAG AGAGTCAGAC 780
AGTGCCACTT GGCCACTTGG GGTAAAGCCA GTGCCAGCAA TAACAGTTTA TCATGCTCAT 840
TAATTTGGGA TTTCAAACA CAAATGAAAA CTCACACCCA CCCACCCCA AGTGCATGTC 900
TCCATCACTT AAAAAGTAAG TTCCATTTGA AAATATCCTT TCTTTTTTTT TTCTTCCTAT 960
TTTTGTTTGT TTATACAAAT ATCTGATTTG CAAGAAAAAG TGCATGGGAG GGGTTTTAGT 1020
GGTTTAATGA ATTTTAAATT AAGAAAGGGT AGTTTGGTAG TCTACTTAAA AATGTTTCTG 1080
GGAAATTCAC TAGAAACATT AACCAATAGG ATTTTGGTGA GCTTAGCTTC TGTATTCCTA 1140
CTGCCGCCCA GAAAAGGGGC AGGGCTCTGC AGCCGCCAGG ACAGACGAGC ACCCATGCC 1200
TATACCTCCC TCCCCGAGCT AAGTCCCAGG GCATCTGGGC CTTGCCTGGA GACTGGGCTA 1260
GCTCTGTAGG CTCGGAGAGC CTGGGGAGGG TGCCAACCCC ACCTCTAGTA TTTTGGGAGA 1320
TAGGGAAAGT GAACCGACTT CCCCTTCCCA TACCCCTCAG GGTGGTTCCC TACCAGCCAG 1380
GCTTACTACT TCTAGAAGAA AGCAGAGTGC CAGGGAGTGA GATTGCATCC CTGGGCTTAG 1440
AAGTGACGGA GAGAAGACTT GTTTAGTATT TTGCCATCAG CACAAGGAAA ACCAGGAGAG 1500
AGTCTGCCTC CAGGACTCTG AGCCTTCTGC CTCGTATGTT CAGAAGGTGG ATAGGTCTTC 1560

```

PF-0459 US

CCACTCCAGC ATGGCTTGAA CTCTTAGGGG TCTGCAGTGC TCCATCTCCA TTGGTGGCCC 1620  
CAGCTCAGTA ACTATACCTG GTACATTTCC TGTGTGCAAT CAGTACCTTG AAGGCAGAAC 1680  
ATTCTGAATA AAGTTGGAAG AAGAACAGCT TTGCTTTGCA AAGATTGATG ACAGACTGGT 1740  
TCCTCAGAGG CCTAGGCTAC CCGTCACCCC TTTTTCAGAG GCGAGGGCCT GGAATGAAGG 1800  
CAGTTTATCC TCTGTCCCTG GAGCCTGGGG TTTGCTTTGG CTCCTTGAGG TGGAAGAGAC 1860  
TAAGAGGGCA GCTGCCCAGA GCAGCTGTGT GTACCTGGCT CCTCTCAGGC TTCCTGATCC 1920  
CTTCCATTGC ACTGCGCCTT ATCCCTCAGC CAGCCAGACA GCCTCCCTGC TCCTGACCAG 1980  
CAGATACGTT TCGGAGTGGT TGGTGTGGTT TTTGTGATGA GGGCAGCACA TGGTGGCCAA 2040  
GGTGGGCAAA GCTGAGTCTC ACAAGGCTCA AATCCCTTCG GTTGGGNTCC CCTTGTGGGG 2100

(2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2458 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PANCNOT08
- (B) CLONE: 1438165

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95 :

GCGGGCGGAG ATGTAGACCC GGTAGTGTG TGCCTTGTGG TGACAACTGG CGGCAGCGCG 60  
CCGCGGGCCC GAGACTTAGT CTCGGGCCGC CATGGCCAGC GTCCACGAGA GCCTCTACTT 120  
CAATCCCATG ATGACCAATG GGGTTGTGCA CGCCAATGTG TTCGGCATCA AGGACTGGGT 180  
GACGCCGTAC AAGATCGCGG TGCTGGTGCT GCTGAACGAG ATGAGCCGCA CAGGCGAGGG 240  
CGCCGTCAGC CTCATGGAGC GCGGAGGCT CAACCAGCTG CTCCTGCCCC TGCTGCAGGG 300  
CCCAGATATT AACTGTCAA AACTTTACAA GTTAATTGAA GAGTCTTGTC CACAGCTGGC 360  
AAATTCAGTG CAGATCAGAA TCAAATGAT GGCTGAAGGC GAGTTGAAGG ATATGGAACA 420  
GTTTTTTGAT GACCTTTCAG ATTCTTCTC TGGAAGTGA CCAGAGGTTT ACAAACAAG 480  
TGTAGTAGGT TTGTTTCTGC GTCACATGAT CTTGGCCTAC AGTAAGCTTT CTTTCAGCCA 540  
AGTGTTTAAA CTGTACACTG CCCTTCAGCA GTACTTCCAG AATGGTGAGA AAAAGACAGT 600  
GGAGGATGCT GATATGGAAC TGACCAGTAG AGATGAGGGT GAAAGAAAAA TGGAAGAAAG 660  
AGAACTTGAT GTATCTGTAA GAGAAGAGGA GGTATCTTGC AGTGGGCCTC TGTCCCAAAA 720  
ACAAGCAGAA TTTTTTCTTT CTCAACAGGC TTCTTTGCTA AAGAATGATG AGACTAAGGC 780

PF-0459 US

CCTCACTCCA GCTTCCTTGC AGAAGGAATT AAACAATTTG TTGAAATTTA ATCCTGATTT 840  
TGCTGAAGCG CATTATCTCA GCTACTTAAA CAACCTCCGT GTCCAAGATG TTTTCAGTTC 900  
AACACACAGT CTCCTCCATT ATTTTGATCG TCTGATTCTT ACCGGAGCCG AAAGCAAAAG 960  
TAATGGGGAA GAGGGCTATG GCCGGAGCTT GAGATACGCC GCTCTGAATC TTGCCGCCCT 1020  
GCACTGCCGC TTCGGTCACT ATCAACAGGC AGAGCTCGCC CTGCAGGAGG CAATTAGGAT 1080  
TGCCCAGGAG TCCAACGATC ACGTGTGTCT CCAGCACTGT TTGAGCTGGC TTTATGTGCT 1140  
GGGGCAGAAG AGATCCGATA GCTATGTTCT GCTGGAGCAT TCTGTGAAGA AGGCAGTACA 1200  
TTTTGGGTTA CCGAGAGCTT TTGCTGGGAA GACGGCAAAC AAGCTGATGG ATGCCCTAAA 1260  
GGACTCCGAC CTCCTGCACT GGAAACACAG CCTGTCAGAG CTCATCGATA TCAGCATCGC 1320  
ACAGAAAACG GCCATCTGGA GGCTGTATGG CCGCAGCACC ATGGCACTGC AACAGGCCCA 1380  
GATGTTGCTG AGCATGAACA GCCTGGAGGC GGTGAATGCG GGCGTGCAGC AGAACAACAC 1440  
AGAGTCCTTT GCTGTGCGAC TCTGCCACCT CGCAGAGCTA CACGCGGAGC AGGGCTGTTT 1500  
TGCTGCAGCT TCTGAAGTGT TAAAGCACTT GAAGGAACGA TTTCCGCCTA ATAGTCAGCA 1560  
CGCCCAGTTA TGGATGCTAT GTGATCAAAA AATACAGTTT GACAGAGCAA TGAATGATGG 1620  
CAAATATCAT TTGGCTGATT CACTTGTTAC AGGAATCACA GCTCTCAATA GCATAGAGGG 1680  
TGTTTATAGG AAAGCGGTTG TATTACAAGC TCAGAACCAA ATGTCAGAGG CACATAAGCT 1740  
TTTACAAAAA TTGTTGGTTC ATTGTCAGAA ACTGAAGAAC ACAGAAATGG TGATCAGTGT 1800  
CCTACTGTCC GTGGCAGAGC TGTACTGGCG ATCTTCCTCC CCTACCATCG CGCTGCCCCAT 1860  
GCTCCTGCAG GCTCTGGCCC TCTCCAAGGA GTACCGGTTA CAGTACTTGG CCTCTGAAAC 1920  
AGTGCTGAAC TTGGCTTTTG CGCAGCTCAT TCTTGAATC CCAGAACAGG CCTTAAGTCT 1980  
TCTCCACATG GCCATCGAGC CCATCTTGGC TGACGGGGCT ATCCTGGACA AAGGTCGTGC 2040  
CATGTTCTTA GTGGCCAAGT GCCAGGTGGC TTCAGCAGCT TCCTACGATC AGCCGAAGAA 2100  
AGCAGAAGCT CTGGAGGCTG CCATCGAGAA CCTCAATGAA GCCAAGAACT ATTTTGCAAA 2160  
GGTTGACTGC AAAGAGCGCA TCAGGGACGT CGTTTACTTC CAGGCCAGAC TCTACCATAC 2220  
CCTGGGGAAG ACCCAGGAGA GGAACCGGTG TGCGATGCTC TTCCGGCAGC TGCATCAGGA 2280  
GCTGCCCTCT CATGGGGTAC CCTTGATAAA CCATCTCTAG AGAGGACATC CCTGCTGGGC 2340  
TGCTGTGCAG AGTATAAGAT TTTGGACTTG TTCATGTCCC CTCTCTCCCT ATAAATGATG 2400  
TATTTGTGAC ACCCTATCTT GTCAATAAAC AGCATCTGA TTAAAAAAAA AAAAAAAA 2458

(2) INFORMATION FOR SEQ ID NO: 96:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2900 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: THYRNOT03  
 (B) CLONE: 1440381

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96 :

TGCATGGATG GGATACTGGA TGAATCTTTG CTTGAAACCT GTCCAATTCA GTCACCATTA 60  
 CAAGTTTTTTG CAGGAATGGG TGGACTGGCT CTTATTGCTG AAAGACTACC CATGCTATAT 120  
 CCAGAAGTAA TTCAACAGGT GAGTGCTCCA GTTGTAAACAT CTACCACTCA GGAAAAGCCG 180  
 TATGATAGCG ATCAGTTTGA ATGGGTGACC ATTGAACAGT CAGGGGAGTT AGTTTATGAA 240  
 GCACCAGAAA CTGTTGCGGC TGAACCTCCA CCTATCAAGT CAGCAGTACA GACCATGTCT 300  
 CCCATACCTG CCCATTCTTT GGCTGCTTTT GGATTATTTT TTCGTCTTCC GGGCTATGCG 360  
 GAAGTGCTAC TGAAAGAGAG AAAACATGCC CAGTGCCTTC TTCGATTGGT ATTGGGAGTG 420  
 ACAGATGATG GAGAAGGAAG TCATATTCTT CAATCTCCAT CAGCCAATGT GCTTCCAACC 480  
 CTTCCCTTTCC ACGTCCTTCG TAGCTTGTTT AGCACTACAC CTTTGACAAC TGATGATGGT 540  
 GTACTTCTAA GGCGGATGGC ATTGGAAATT GGAGCCTTAC ACCTCATCTT TGTCTGTCTC 600  
 TCTGCTTTGA GCCACCATTG CCCACGAGTT CCAAACCTCA GCGTGAATCA AACTGAGCCA 660  
 CAGGTGTCAA GCTCTCATAA CCCTACATCA ACAGAAGAAC AACAGTTATA TTGGGCCAAA 720  
 GGGACTGGCT TTGGAACAGG CTCTACAGCT TCTGGGTGGG ATGTGGAACA AGCCTTAACT 780  
 AAGCAAAGGC TGGAAGAGGA ACATGTTACC TGCCTTCTGC AGGTTCTTGC CAGTTACATA 840  
 AATCCCGTCA GTAGTGCGGT AAATGGAGAA GCTCAGTCAT CTCATGAGAC TAGAGGGCAG 900  
 AACAGTAATG CCCTTCCTTC TGTACTTCTC GAGCTTCTCA GTCAGTCCTG CCTCATCCCA 960  
 GCCATGTCAT CTTATCTACG AAATGATTCA GTTCTGGACA TGGCAAGACA TGTGCCACTC 1020  
 TATCGGGCAC TGCTGGAATT GCTTCGGGCC ATTGCTTCTT GTGCTGCCAT GGTGCCCCTA 1080  
 TTGTTGCCCC TTTCTACAGA GAACGGTGAA GAGGAAGAAG AACAGTCAGA ATGTCAAACCT 1140  
 TCTGTTGGTA CATTGTTAGC CAAAATGAAG ACCTGTGTTG ATACCTATAC CAACCGTTTA 1200  
 AGATCTAAAA GGGAAAATGT TAAAACAGGA GTAAAACCAG ATGCGTCTGA TCAAGAACCA 1260  
 GAAGGACTTA CTCTTTTGGT ACCAGACATC CAAAAGACTG CTGAGATAGT TTATGCAGCC 1320  
 ACCACCAGTT TGCGGCAAGC AAATCAGGAA AAAAAGCTGG TGAATACTCC AAGAAGGCGG 1380  
 CTAATGAACC CCAAACCTTT GTCAGTATTA AAGTCACTTG AAGAAAAATA TGTGGCTGTT 1440  
 ATGAAGAAAT TACAGTTTGA TACGTTTGAA ATGGTTTCTG AAGATGAAGA TGGGAAATTG 1500

PF-0459 US

GGATTTAAAG TAAATTACCA CTACATGTCT CAGGTGAAAA ATGCTAATGA TGCGAACAGT 1560  
GCTGCCAGAG CTCGCCGCCT TGCCAGGAA GCTGTGACGC TTTCAACCTC ACTGCCTCTG 1620  
TCTTCATCCT CTAGTGTGTT TGTACGCTGT GATGAGGAGC GACTTGATAT CATGAAGGTT 1680  
CTAATAACTG GTCCAGCGGA CACCCCTTAT GCAAATGGCT GCTTTGAGTT TGATGTGTAT 1740  
TTTCCTCAAG ATTATCCAG TTCACCCCTT CTTGTGAATC TAGAGACAAC TGGTGGTCAT 1800  
AGCGTGCGAT TCAATCCAAA CCTTTATAAT GATGGCAAGG TTTGTTTAAG CATCTTAAAC 1860  
ACGTGGCATG GAAGACCAGA AGAGAAGTGG AATCCTCAGA CCTCAAGCTT TTTGCAAGTG 1920  
TTGGTGTCTG TCCAGTCCCT TATATTAGTA GCTGAGCCTT ATTTTAATGA ACCGGGATAT 1980  
GAACGGTCTA GAGGCACTCC CAGTGGCACA CAGAGTTCTC GAGAATATGA TGGAAACATT 2040  
CGACAAGCAA CAGTTAAGTG GGCAATGCTA GAACAAATCA GAAACCCTTC ACCATGTTTT 2100  
AAAGAGGTAA TACACAAACA TTTTACTTG AAAAGAGTTG AGATAATGGC CCAATGTGAG 2160  
GAGTGGATTG CGGATATCCA GCAGTACAGC AGTGATAAGC GGGTAGGCAG GACTATGTCT 2220  
CACCATGCAG CAGCTCTCAA GCGTCACACT GCTCAGCTCC GCGAAGAGTT GCTGAAACTT 2280  
CCCTGCCCTG AAGGCTTGGA TCCTGACACT GACGATGCCC CAGAGGTGTG CAGAGCCACA 2340  
ACAGGTGCTG AGGAGACTCT AATGCATGAT CAGGTTAAAC CCAGCAGCAG CAAAGAACTC 2400  
CCCAGTGACT TCCAGTTATG AGCTGCATTG ATGTGGACTT CATAGACACA AAGGCTTCGA 2460  
AGCACAAGCC AAATATGTCA ATATTTGTAT GTAAGAACT AATTATGTAA TAGGTAATGA 2520  
AACTGAACT ATACTATGCC CTTAAGGAGA TCCAGTTTAA TTCAAGGTGA TCTTTTATTT 2580  
ACCTGTACAG GAGTGTAAC TTTTTGTGC TTTTATTTT CAATTGTGAG AACCCTGAT 2640  
TGGTATGTTT AACAAATTTG TGTATACAAA GAAATGGATA AATCACTGCT ATATAAGGGA 2700  
AACTACCTTA GGAAAGAATG TTTACTGAAT GTTTATTTTA TTTTATTTT TTTTACTAT 2760  
AGAGTGAGGG GTTGTTAACA AAGAATATAT ATTGGTCGTT CTTACAATA CTATTTAAAG 2820  
TCAGCAACTT TTCCTGAAT TTGATAGATT TTATGTTTG GGGTACGAGC TTGTAAAGCT 2880  
CGGGTGCCTN ATGAGTGACC 2900

(2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1310 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: LUNGNOT14  
(B) CLONE: 1510839

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97 :

```

CCGCTGAGAT GTACGAACTT CCGGTTCTCC GGGCAGCTGC CACTGCTGTA GCTTCTGCCA 60
CCTGCCACGA CCGGGCCTCT CCCTGGCGTT TGGTCACCTC TGCTTCATTC TCCACCGCGC 120
CTATGGTCCC TCTTGGAGCC AGCGTGGCGG GCCTGGCGGC TCCCGGGTGG TGAGAGAGCG 180
GTCCGGGAAC GATGAAGGCC TCGCAGTGCT GCTGCTGTCT CAGCCACCTC TTGGCTTCCG 240
TCCTCCTCCT GCTGTTGCTG CCTGAACTAA GCGGGCCCCCT GGCAGTCCTG CTGCAGGCAG 300
CCGAGGCCGC GCCAGGTCTT GGGCCTCCTG ACCCTAGACC ACGGACATTA CCGCCGCTGC 360
CACCGGGCCC TACCCCTGCC CAGCAGCCGG GCCGTGGTCT GGCTGAAGCT GCGGGGCCGC 420
GGGGCTCCGA GGGAGGCAAT GGCAGCAACC CTGTGGCCGG GCTTGAGACG GACGATCACG 480
GAGGGAAGGC CGGGGAAGGC TCGGTGGGTG GCGGCCCTTG TGTGAGCCCC AACCCCTGGC 540
ACAAGCCCAT GACCCAGCGG GCCCTGACCG TGTGATGGT GGTGAGCGGC GCGGTGCTGG 600
TGTACTIONG GGTGAGGACG GTCAGGATGA GAAGAAGAAA CCGAAAGACT AGGAGATATG 660
GAGTTTTGGA CACTAACATA GAAAATATGG AATTGACACC TTTAGAACAG GATGATGAGG 720
ATGATGACAA CACGTTGTTT GATGCCAATC ATCCTCGAAG AAGAGAATGT GCCTTTTGAT 780
GAAAGAACTT TATCTTTCTA CAATGAAGAG TGAATTTCT ATGTTTAAGG AATAAGAAGC 840
CACTATATCA ATGTTGGGGG GGTATTTAAG TTACATATAT TTTAACAACC TTTAATTTGC 900
TGTTGCAATA AATACCGTAT CCTTTTATTA TATCTTTATA TGTATAGAAG TACTCTATTA 960
ATGGGCTCAG AGATGTTGGG GATAAAGTAT ACTGTAATAA TTTATCTGTT TGAAATTAC 1020
TATAAAACGG TGTTTTCTGA TCGGTTTTTG TTTCTGCTT ACCATATGAT TGTAATTGT 1080
TTTATGTATT AATCAGTTAA TGCTAATTAT TTTGCTGAT GTCATATGTT AAAGAGCTAT 1140
AAATTCCAAC AACCAACTGG TGTGTAAAAA TAATTTAAAA TTTCTTTAC TGAAAGGTAT 1200
TTCCCATTTT TGTGGGAAA AGAAGCCAAA TTTATTACTT TGTGTTGGGG TTTTAAAAAT 1260
ATTAAGAAAT GTCTAAGTTA TTGTTTGCAA AACAATAAAT ATGATTTTAG 1310

```

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2272 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: SPLNNOT04  
 (B) CLONE: 1534876

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98 :

CCATGCTCCA GGCATACAGA TGTGGTTTCT CGGCTGCACC GGGCCAGGCT GCGGGTGTGC 60  
 AGGCGTCTGC AAAGTTGTGC CATGTATCAG CACAGGCTTT GAGACGTCTG GACCCTGTCC 120  
 TTCCTCCCGT GAGGGGTTCT TGTTCTTTCT GACTCAGGTG ACTTTTCAGC CCTTCCAATT 180  
 CCCCTCTTTT TCTGCCCTCC CCTCCAATC AGCCAACCCA GGTGTGGGCA GTCAGGGAGG 240  
 GAGGGAGTGT CCCACCACGT TCTCAGGGCA GCCCTTGACT CCTAAGCCCC TTCCTCCTTC 300  
 CATTCTGCAT CCCCTCCCCA TCCAACCTAA ATGCCACAG CTGGGGCTGA GCTGTATTCC 360  
 TGTGGAGGGA CCTCTGCCGT GCCTCTCTGA GGTGAGGCTG TGCTGTGTGA TGGGCAGGCT 420  
 TTGCCCCAGC CCACCCCTGG CAAGGTGCAC TTGTTTTCTG GTTTGTACAA GGTGTCCTGG 480  
 GGGCCCGTCG CTTCCCTGCC AGTGAGGAGT GACTTCTCCC TCTCTCCAG TCCTGTAGGG 540  
 GAGACAAAAC CAGATTGGGG GGGCCAAGGG GAGCATGGAA AAGGCCGGCT CCCCTGTCTT 600  
 TCCTTGCGTG TCAGAGTCAG GGTAACACAC ACCAAGAGTG GAGTGCGGCC AGCAAGTTTG 660  
 AGACCTGCCC GCCCTCCTCG CAGCTCTGCT CTGTGTCCTC AGGAAGTCAC AGAGTCTACT 720  
 GAGGCAAGGA GAGGGTGATT CTTTCCCCAA ATCCCTTCTT CCCTGGTTCC CAAACCAAAG 780  
 ACAGCCTGCA GCCCTTTCTG CATGGGGTGC TCTGTTGACA GGCTTCCCAG ATCCCTGAGT 840  
 CTCTCTTTCC TTCCTCCTCG ATCTTTAGTT GTCCACGGTC AATTCAGTGC TTCCATTGGG 900  
 GGACAGTCCC CTCCGGGATG ACCTGATTCA CCTCCAGCCC AGGGAATGGA ATCTAGAGGA 960  
 ATACGTGGGG TGGGTCTGGA CAAGGAGCGG CAGGAATCAC CACCCATCTC CAGCTGTGGA 1020  
 GCCCTGTGGA GGGGAAGGGG AAGCTTGGGG TTCAGAGGGA ACTCTTCCAG GAGAGGGGTG 1080  
 CCCAGCGGAG GTAAAGATGA TAGAGGGTTG TGGGGGTCT CTAGTTGAAT GTTTTGGCCC 1140  
 ATGACTTTGG AACATGGCTG GCAGCTTCCA GCAGAAGTCA CGCTCCCCAT CCCCAGGGG 1200  
 ACATAGGACC TTTTCTCTGC TTCCTGGTCA CTTTCAAAGA ACTATTTGCG CAATCTGTGG 1260  
 GTCTGTGGAT TCACGGGGCT TTCTGTGTGG GTGCTGCAGT TGCTTTTGTC TGCAGCAGCA 1320  
 GGACACATCT TTCCTCTTAC TCAGCCCTTT ATGGCCCATG GGGAACTCCG TGGCTCAGGG 1380  
 AGAGCTGAAC TCCAGGGGTG TGACCTGGGA CAGGTGGGCC TGAGGTGCCC AGCTCAGGGC 1440  
 AGCCAGGTGG CTCATGGGCT GTAGTGAGCC AGCTCCCTGG GGGAAAAGGC TGTGGGCCGT 1500  
 TAGGACCATC CTCCAGGACA GGTGACCTCT ATGAGGTCAC CTACGGCTGT GGCCGTGCAG 1560  
 GCCTCCTTCC AGCCCAGAGT GGGCCAGTAG AGCAAGGCAG ACAGTGACCT CCACCCCCGC 1620  
 AGCCCTCTTA AAAGGCCAGT ACTCTTGGGG GTGGGGGGAG GGTTTAGAAA GCATTTGCCC 1680



PF-0459 US

ATCTGCCTTT CTTTCCCCCA GCCCCACCC GCTTTGAATG TAGAGACCCG TGGGCACTTT 1740  
TCCTTTTGTG GTGGGGGGTG CGGAGGAGGT ACCCCCACCC CTGGCACAGC CGCCTGGAAT 1800  
GCAGGACTGT CACTGCTGTT CGGGTGATGA CCTCGTTGCC AAGCTCCTCC TGTCCCCTTG 1860  
TTCTGGGGGC AGGCGCTGTG CTTCTGTGAG GTGGTTTAGC TTTTGCTTTC GAAGTGGCCA 1920  
GCTGCGGCCA CCAGGTCTCA GCACAAGAGC GCTTCCTTTG CACAGAATGA GCTTCGAGCT 1980  
TTGTTTCTAG TAAATGAATG TATCTGGGAG GGGTCGGGGG CACGAGTTGA TTCCAAGCAC 2040  
ATGCCTTTGC TGAGTGTGTG TGTGCTGGGA GAGTCAGAGT GGATGTAGAG CGCGGTTTTA 2100  
TTTTTGTACT GACATTGGTA AGAGACTGTA TAGCATCTAT TTATTTAGAT GATTATCTG 2160  
GTAAATGAGG CAAAAAATT ATTAATAA CATTAAAGAT GATTAAAAA AAAGACCAAA 2220  
AAACCAAGAA ACCCAAAGCC CAAGAATGCG CGTAGCATCC AAAAAAAAAA GG 2272

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1060 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: SPLNNOT04
- (B) CLONE: 1559131

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99 :

GTCAACTTAG CGAGCGCAAC AGGCTGCCGC TGAGGAGCTG GAGCTGGTGG GGACTGGGCC 60  
GCAATGGACA AGCTGAAGAA GGTGCTGAGC GGGCAGGACA CGGAGGACCG GAGCGGCCTG 120  
TCCGAGGTTG TTGAGGCATC TTCATTAAGC TGGAGTACCA GGATAAAAGG CTTTATTGCG 180  
TGTTTTGCTA TAGGAATTCT CTGCTCACTG CTGGGTACTG TTCTGCTGTG GGTGCCCAGG 240  
AAGGGACTAC ACCTCTTCGC AGTGTTTTAT ACCTTTGGTA ATATCGCATC AATTGGGAGT 300  
ACCATCTTCC TCATGGGACC AGTGAAACAG CTGAAGCGAA TGTTTGAGCC TACTCGTTTG 360  
ATTGCAACTA TCATGGTGCT GTTGTGTTTT GCACTTACCC TGTGTTCTGC CTTTTGGTGG 420  
CATAACAAGG GACTTGCACT TATCTTCTGC ATTTTGCACT CTTTGGCATT GACGTGGTAC 480  
AGCCTTTCCT TCATACCATT TGCAAGGGAT GCTGTGAAGA AGTGTTTTGC CGTGTGTCTT 540  
GCATAATTCA TGGCCAGTTT TATGAAGCTT TGGAAGGCAC TATGGACAGA AGCTGGTGGA 600  
CAGTTTTGTA ACTATCTTCG AAACCTCTGT CTTACAGACA TGTGCCTTTT ATCTTGCAGC 660  
AATGTGTTGC TTGTGATTGC AACATTTGAG GGTTACTTTT GGAAGCAACA ATACATTCTC 720

PF-0459 US

GAACCTGAAT GTCAGTAGCA CAGGATGAGA AGTGGGTTCT GTATCTTGTG GAGTGAATC 780  
TTCCTCATGT ACCTGTTTCC TCTCTGGATG TTGTCCCACT GAATTCCCAT GAATACAAAC 840  
CTATTCAGCA ACAGCACATA AGCCTTGGGT GCAAGTGATT CCCAGGTGGC AAAAGGCAGC 900  
CCCATCAGAG ATCACGGGAG CAACAGTAAG GGACAGAGTT TTGGGGTCCA CTTGTCCCTC 960  
AGCATGGAAG CCATCACCGT GGTCTGTCAT AGAGTGAGTC TGCTTCTACT CTGGCATCTG 1020  
AGAACAAGTG ACTCTGCTTT AGACAAGCCC CTGGAGAGGG 1060

(2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 543 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BLADNOT03
- (B) CLONE: 1601473

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100 :

GCTCACAGTA GCCCGGCGGC CAGGGCAATC CGACCACATT TCACTCTCAC CGCTGTAGGA 60  
ATCCAGATGC AGGCCAAGTA CAGCAGCACA AGGGACATGC TGGATGATGA TGGGGACACC 120  
ACCATGAGCC TGCATTCTCA AGCCTCTGCC ACAACTCGGC ATCCAGAGCC CCGGCGCACA 180  
GAGCACAGGG CTCCCTCTTC AACGTGGCGA CCAGTGGCCC TGACCCTGCT GACTTTGTGC 240  
TTGGTGCTGC TGATAGGGCT GGCAGCCCTG GGGCTTTTGT GTAAGTCTGC GCTCTGACCT 300  
GGGGGAGGAT CCTGGTTCCA AGTTTTTCAG TACTACCAGC TCTCCAATAC TGGTCAAGAC 360  
ACCATTTCTC AAATGGAAGA AAGATTAGGA AATACGTCCC AAGAGTTGCA ATCTCTTCAA 420  
GTCCAGAATA TAAAGCTTGC AGGAAGTCTG CAGCATGTGG CTGAAAAACT CTGTCGTGAG 480  
CTGTATAACA AAGCTGGAGC ACACAGGTGC AGCCCTTGTA CAGAACAATG GAAATGGCAT 540  
GGA 543

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2281 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

PF-0459 US

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: BRAITUT12  
(B) CLONE: 1615809

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101 :

AGCTGGCTCA CCTTCCAGAT TCACCTGCAG GAGCTGCTGC AGTACAAGAG GCAGAATCCA 60  
GCTCAGTTCT GCGTTCGAGT CTGCTCTGGC TGTGCTGTGT TGGCTGTGTT GGGACACTAT 120  
GTTCCAGGGA TTATGATTTC CTACATTGTC TTGTTGAGTA TCCTGCTGTG GCCCCTGGTG 180  
GTTTATCATG AGCTGATCCA GAGGATGTAC ACTCGCCTGG AGCCCCTGCT CATGCAGCTG 240  
GACTACAGCA TGAAGGCAGA AGCCAATGCC CTGCATCACA AACACGACAA GAGGAAGCGT 300  
CAGGGGAAGA ATGCACCCCC AGGAGGTGAT GAGCCACTGG CAGAGACAGA GAGTGAAAGC 360  
GAGGCAGAGC TGGCTGGCTT CTCCCCAGTG GTGGATGTGA AGAAAACAGC ATTGGCCTTG 420  
GCCATTACAG ACTCAGAGCT GTCAGATGAG GAGGCTTCTA TCTTGAGAG TGGTGGCTTC 480  
TCCGTATCCC GGGCCACAAC TCCGCAGCTG ACTGATGTCT CCGAGGATTT GGACCAGCAG 540  
AGCCTGCCAA GTGAACCAGA GGAGACCCTA AGCCGGGACC TAGGGGAGGG AGAGGAGGGA 600  
GAGCTGGCCC CTCCCGAAGA CCTACTAGGC CGTCTCAAG CTCTGTCAAG GCAAGCCCTG 660  
GACTCGGAGG AAGAGGAAGA GGATGTGGCA GCTAAGGAAA CCTTGTTGCG GCTCTCATCC 720  
CCCCCTCACT TTGTGAACAC GCACTTCAAT GGGGCAGGGT CCCCCAAGA TGGAGTGAAA 780  
TGCTCCCCTG GAGGACCACT GGAGACACTG AGCCCCGAGA CAGTGAGTGG TGGCCTCACT 840  
GCTCTGCCCC GCACCCTGTC ACCTCCACTT TGCTTGTGTT GAAGTGACCC AGCCCCCTCC 900  
CCTTCCATTC TCCCACCTGT TCCCCAGGAC TCACCCCAGC CCCTGCCTGC CCCTGAGGAA 960  
GAAGAGGCAC TCACCACTGA GGACTTTGAG TTGCTGGATC AGGGGGAGCT GGAGCAGCTG 1020  
AATGCAGAGC TGGGCTTGGA GCCAGAGACA CCGCCAAAAC CCCCTGATGC TCCACCCCTG 1080  
GGGCCCCGACA TCCATTCTCT GGTACAGTCA GACCAAGAAG CTCAGGCCGT GGCAGAGCCA 1140  
TGAGCCAGCC GTTGAGGAAG GAGCTGCAGG CACAGTAGGG CTTCTTGGCT AGGAGTGTG 1200  
CTGTTTCCTC CTTTGCCTAC CACTCTGGGG TGGGGCAGTG TGTGGGGAAG CTGGCTGTG 1260  
GATGGTAGCT ATTCCACCCT CTGCCTGCCT GCCTGCCTGC TGTCTGGGC ATGGTGCAGT 1320  
ACCTGTGCCT AGGATTGGTT TTAAATTTGT AAATAATTTT CCATTTGGGT TAGTGGATGT 1380  
GAACAGGGCT AGGGAAGTCC TTCCCACAGC CTGCGCTTGC CTCCTGCCT CATCTCTATT 1440  
CTCATTCAC TATGCCCCAA GCCCTGGTGG TCTGGCCCTT TCTTTTCTCT CCTATCCTCA 1500  
GGGACCTGTG CTGCTCTGCC CTCATGTCCC ACTTGGTTGT TTAGTTGAGG CACTTTATAA 1560  
TTTTTCTCTT GTCTGTGTT CCTTTCTGCT TTATTTCCCT GCTGTGTCCT GTCCTTAGCA 1620  
GCTCAACCCC ATCCTTTGCC AGCTCCTCCT ATCCCGTGGG CACTGGCCAA GCTTTAGGGA 1680

PF-0459 US

GGCTCCTGGT CTGGGAAGTA AAGACTAAAC CTGGGGCAGT GGGTCAGGCC AGTAGTTACA 1740  
CTCTTAGGTC ACTGTAGTCT GTGTAACCTT CACTGCATCC TTGCCCCATT CAGCCCGGCC 1800  
TTTCATGATG CAGGAGAGCA GGGATCCCGC AGTACATGGC GCCAGCACTG GAGTTGGTGA 1860  
GCATGTGCTC TCTCTTGAGA TTAGGAGCTT CCTTACTGCT CCTCTGGGTG ATCCAAGTGT 1920  
AGTGGGACCC CCTACTAGGG TCAGGAAGTG GACACTAACA TCTGTGCAGG TGTTGACTTG 1980  
AAAAATAAAG TGTTGATTGG CTAGAACTGC TGCCTCCCTG ACTGTGAGCT GCCTTCCACA 2040  
CCCTGCACTG CACTGTGTTC TCTCCTCACC CTTAACCTGC TTTACTCCAG TCTGTTCTGG 2100  
CTGTTTATTA CCTTGTTGCA AAACAGGGCC GAAGCAAGGA TTACCTTGAC AACCTAGCT 2160  
TCTCCTTAGC CATCTTCCTT GACAGTGTGA TCTGTTTAGT GAGATTTAGC ATGTGTGAAT 2220  
AAAGTATATG CAGGAGGAAA TTGCTTTGTC TTCCCAATCG GTAGAAATTC GAGACCTAGC 2280  
C 2281

(2) INFORMATION FOR SEQ ID NO: 102:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 992 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: COLNNOT19  
(B) CLONE: 1634813

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102 :

GACAGCTTGG CCTACAGCCC GGCGGGCATC AGCTCCCTTG ACCCAGTGGG TATCGGTGGC 60  
CCCCTTATTC GTCCAGGTGC CCAGGGAGGA GGACCCGCCT GCAGCATGAA CCTGTGGCTC 120  
CTGGCCTGCC TGGTGGCCGG CTTCTTGGA GCCTGGGCCC CCGCTGTCCA CGCCCAAGGT 180  
GTCTTTGAGG ACTGCTGCCT GGCCTACCAC TACCCCATTT GGTGGGCTGT GCTCCGGCGC 240  
GCCTGGACTT ACCGGATCCA GGAGGTGAGC GGGAGCTGCA ATCTGCCTGC TGCGATATTC 300  
TACCTCCCCA AGAGACACAG GAAGGTGTGT GGGAACCCCA AAAGCAGGGA GGTGCAGAGA 360  
GCCATGAAGC TCCTGGATGC TCGAAATAAG GTTTTTGCAA AGCTCCGCCA CAACACGCAG 420  
ACCTTCCAAG CAGGCCCTCA TGCTGTAAAG AAGTTGAGTT CTGGAAACTC CAAGTTATCA 480  
TCATCCAAGT TTAGCAATCC CATCAGCAGC AGCAAGAGGA ATGTCTCCCT CCTGATATCA 540  
GCTAATTCAG GACTGTGAGC CGGCTCATTT CTGGGCTCCA TCGGCACAGG AGGGGCCGGA 600  
TCTTTCTCCG ATAAAACCGT CGCCCTACAG ACCCAGCTGT CCCCACGCCT CTGTCTTTTG 660

PF-0459 US

GGTCAAGTCT TAATCCCTGC ACCTGAGTTG GTCCTCCCTC TGCACCCCCA CCACCTCCTG 720  
CCCGTCTGGC AACTGGAAAG AGGGAGTTGG CCTGATTTTA AGCCTTTTGC CGCTCCGGGG 780  
ACCAGCAGCA ATCCTGGGCA GCCAGTGGCT CTTGTAGAGA AGACTTAGGA TACCTCTCTC 840  
ACTTTCTGTT TCTTGCCGTC CACCCCGGGC CATGCCAGTG TGTCCCTCTG GGTCCCTCCA 900  
AAACTCTGGT CAGTTCAAGG ATGCCCTCC CAGGCTATGC TTTTCTATAA CTTTTAAATA 960  
AACCTTGGGG GTTGATGGAG TCAAAAAAAA AA 992

(2) INFORMATION FOR SEQ ID NO: 103:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1554 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: UTRSNOT06  
(B) CLONE: 1638407

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103 :

TCGCCCAGGA GTCATCGGAC GCCAGAATCT GTGTCTCCAG AACGCTATAG CTATGGCACC 60  
TCCAGCTCTT CAAAGAGGAC AGAGGGTAGC TGCCGTCGCC GTCGGCAGTC AAGCAGTTCT 120  
GCAAATTCTC AGCAGGGTCA GTGGGAGACA GGCTCCCCCC CAACCAAGCG GCAGCGGCGG 180  
AGTCGGGGCC GGCCAGTGG TGGTGCCAGA CGGCGGCGGA GAGGGGCCCC AGCCGCACCC 240  
CAGCAGCAGT CAGAGCCCGC CAGACCTTCC TCTGAAGGCA GGTGACACTG TGATGGGGAA 300  
ACAGGCTCAG AGAGACATCC GGCTCCGGGT TCGAGCAGAG TACTGCGAGC ATGGGCCAGC 360  
CTTGAGCAG GGCCTGGCAT CCCGGCGGCC CCAGGCGCTG GCGCGGCAGC TGGACGTGTT 420  
TGGGCAGGCC ACCGCAGTGC TGCCTCAAG GGACCTGGGC TCTGTGTTTT GTGACATCAA 480  
GTTCTCAGAG CTCTCCTATC TGGACGCCTT CTGGGGCGAC TACCTGAGTG GCGCCCTGCT 540  
GCAGGCCCTG CGGGGCGTGT TCCTGACTGA GGCCCTGCGA GAGGCTGTGG GCCGGGAGGC 600  
TGTTGCGCTG CTGGTCAGTG TGGATGAGGC TGACTATGAG GCTGGCCGGC GCCGCCTGTT 660  
GCTGATGGCG GAGGAAGGGG GGCGGCGCCC GACAGAGGCC TCCTGATCCA GGA CTGGCAG 720  
GATTGATCCC ACCTCCAAGT CTCCGGGCCA CCTTCTCCTG GGAGGACGAC CATCTCTACC 780  
CCTAGAGGAC TGTCACCTA GCATCTTTGA GGACTGCGAC AGGACCGGGA CAGCAGGCCC 840  
CTTGACAGCC CCTCCACAG GATGTGGGCT CTGAGGCCTA AACCATTTC AGCTGAGTTT 900  
CCTTCCCAGA CTCCTCCTAC CCCCAGGTGT GCCCCCTTAG CCTCCGGAGG CGGGGGCTGG 960

PF-0459 US

GCCTGTATCT CAGAAGGGAG GGGCACAGCT ACACACTCAC CAAAGGCCCC CCTGCACATT 1020  
GTATCTCTGA TCTTGGGCTG TCTGCACTGT CACAGGTGCA CACACTCGCT CATGCTCACA 1080  
CTGCCCCTGC TGAGATCTTC CCTGGGCCTC TGCCCTGGCC TGCTTCCCAG CACACACTTC 1140  
TTTGGCCTAA GGGCTTCTCT CTCAGGACCT CTAATTTGAC CACAACCAAC CTGGGCTTCA 1200  
GCCACATCAG TGGGCACTGG AGCTGGGGTG CACATGGGGC CTGCTCACCT TGCCCACACA 1260  
TCTCCAGCCA GCCAGGGCCC TGCCAGCTT CAATTTACAG ACCTGACTCT CCTCACCTTC 1320  
CCCCCTGCTG TCCAGAGCTG AACATAGACT TGCACTTGGA GTTCACCTGG AGTGTCACAT 1380  
GGGAGTGTTA TGGCAGCATC ATACCAAGGC CTACTGTTGC ACATGGGGCC AAAACCAGTA 1440  
AACAGCCACC TTCTTGAAA GGAATGCAA AGGCTTTGGG GGTGATGGAA AAGACCTTTT 1500  
ACAAATGATA CCAATTAAAC TGCCCTGGAA AGGGCATAGG TGGGAAAAAA AAAA 1554

(2) INFORMATION FOR SEQ ID NO: 104:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1802 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: PROSTUT08  
(B) CLONE: 1653112

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104 :

GTCGCCGGGC TTGCGATGAA CTTCCGGCTG TCAAGCTCCC GGCCGGGCTG ACTCAAGCGG 60  
AGGCGCGCGG AACAGTCGCC GAGGCGATTC CCGCCCAGGC TCCTGTAACC GCCAGGCAGC 120  
GGCCCCGCCA TGTCCCAGCC CCGGACCCCA GAGCAGGCAC TGGATACACC GGGGGACTGC 180  
CCCCCAGGCA GGAGAGACGA GGACGCTGGG GAGGGGATCC AGTGCTCCCA ACGCATGCTC 240  
AGCTTCAGTG ACGCCCTGCT GTCCATCATC GCCACCGTCA TGATCCTGCC TGTGACCCAC 300  
ACGGAGATCT CCCCAGAACA GCAGTTCGAC AGAAGTGTAC AGAGGCTTCT GGCAACACGG 360  
ATTGCCGTCT ACCTGATGAC CTTTCTCATC GTGACAGTGG CCTGGGCAGC ACACACAAGG 420  
TTGTTCCAAG TTGTTGGGAA AACAGACGAC ACACTTGCCC TGCTCAACCT GGCCTGCATG 480  
ATGACCATCA CCTTCCTGCC TTACACGTTT TCGTTAATGG TGACCTTCCC TGATGTGCCT 540  
CTGGGCATCT TCTTGTCTG TGTGTGTGTG ATCGCCATCG GGGTCGTGCA GGCCTGATT 600  
GTGGGGTACG CATTCCACTT CCCGCACCTG CTGAGCCCGC AGATCCAGCG CTCTGCCCAC 660  
AGGGCTCTGT ACCGACGACA CGTCCTGGGC ATCGTCCTCC AAGGCCCGGC CCTGTGCTTT 720

PF-0459 US

GCAGCGGCCA TCTTCTCTCT CTTCTTTGTC CCCTTGTCTT ACCTGCTGAT GGTGACTGTC 780  
ATCCTCCTCC CCTATGTCAG CAAGGTCAAC GGCTGGTGCA GAGACAGGCT CCTGGGCCAC 840  
AGGGAGCCCT CGGCTCACCC AGTGAAGTC TTCTCGTTTG ACCTCCACGA GCCACTCAGC 900  
AAGGAGCGCG TGAAGCCTT CAGCGACGGA GTCTACGCCA TCGTGGCCAC GCTTCTCATC 960  
CTGGACATCT GCGAAGACAA CGTCCCGGAC CCCAAGGATG TGAAGGAGAG GTTCAGCGGC 1020  
AGCCTCGTGG CCGCCCTGAG TGCAGCCGGG CCGCGCTTCC TGGCGTACTT CGGCTCCTTC 1080  
GCCACAGTGG GACTGCTGTG GTTCGCCCAC CACTCACTCT TCCTGCATGT GCGCAAGGCC 1140  
ACGCGGGCCA TGGGGCTGCT GAACACGCTC TCGCTGGCCT TCGTGGGTGG CCTCCCACTA 1200  
GCCTACCAGC AGACCTCGGC CTTGCCCCGG CAGCCCCGCG ATGAGCTGGA GCGCGTGCCT 1260  
GTCAGCTGCA CCATCATCTT CCTGGCCAGC ATCTTCCAGC TGGCCATGTG GACCACGGCG 1320  
CTGCTGCACC AGGCGGAGAC GCTGCAGCCC TCGGTGTGGT TTGGCGGCCG GGAGCATGTG 1380  
CTCATGTTTCG CCAAGCTGGC GCTGTACCCC TGTGCCAGCC TGCTGGCCTT CGCCTCCACC 1440  
TGCTGCTGA GCAGGTTTCAG TGTGGGCATC TTCCACCTCA TGCAGATCGC CGTGCCCTGC 1500  
GCCTTCCTGT TGCTGCGCCT GCTCGTGGGC CTGGCCCTGG CCACCCTGCG GGTCTGCGG 1560  
GGCCTCGCCC GGCCCGAACA CCCCCGCCA GCCCCACGG GCCAGGACGA CCCACAGTCC 1620  
CAGCTCCTCC CTGCCCCCTG CTAGCAGCCA CAGAGCCCAC TCCCAGCCGT CCTCACCAGA 1680  
GATGGACCAG GGAGGACAGG ATGCTGGGCA GGGGAAGCCA AGTCACGGG AGGCCGCAGT 1740  
GGTTCTTGCG TGGCCTGGTT TTATTTTCAT TGTGAAATAT CATGCTCTTA TTTCAGTCCT 1800  
CA 1802

(2) INFORMATION FOR SEQ ID NO: 105:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1395 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: BRSTNOT09  
(B) CLONE: 1664634

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105 :

GTACCTCGGC TTATTTTCATA AACAGGTACT GAAGGAAGCA GAGGCATGTG GAGGACTTCC 60  
CCACCTCGTG CAGCTATTTG GGCCGTGGCA TCTGAAATTT CTTATTTTCAG AGTCACCCCT 120  
TTGATGACCT TGGCAGTGAA CTGCAGTCAT CTGTTTAGGC CTTTCCATGG CCCACGTCAA 180

PF-0459 US

TGCCGGTATT TCTGTTTGTT GCACATTGTA TTTCTTTGTT GTTGGCATT AGAAGGCCCT 240  
CGAGCCGCAC TGAGGGACTG AGCCTGGTGT ATATGGCAGC AAGACTGGAT GGTGGCTTTG 300  
CAGCAGTCTC CAGAGCATTC CATGAGATCC GGGCTCGAAA TCCAGCATTT CAGCCACAAA 360  
CTTTGATGGA CTTTGGCTCA GGTACTGGTT CTGTCACCTG GGCTGCTCAC AGTATTTGGG 420  
GCCAGAGCCT ACGTGAATAT ATGTGTGTGG ACAGATCAGC TGCCATGTTG GTTTTGGCAG 480  
AAAAACTACT GACAGGTGGT TCAGAATCTG GGGAGCCTTA TATTCCAGGT GTCTTTTCA 540  
GACAGTTTCT ACCTGTATCA CCCAAGGTGC AGTTTGATGT AGTAGTGTCA GCTTTTTCT 600  
TAAGTGACCA GCTACTGACA TTTATACTTT CGTGTAATTC AAGTCTTCTG CATATTTTCC 660  
CCTTTTGTGA ACAGGTACTG GTGGAGAATG GAACAAAAGC TGGGCACAGC CTTCTCATGG 720  
ATGCCAGGGA TCTGGTCCTT AAGGGAAAAG AGAAGTCACC TTTGGACCCT CGACCTGGTT 780  
TTGTCTTTGC CCCGTGTCCC CATGAACTCC CTTGTCCCCA GTTGACCAAC CTGGCCTGTA 840  
GCTTCTCACA GCGGTACCAT CCCATCCCCT TCAGCTGGAA CAAGAAACCA AAGGAAGAAA 900  
AGTTCTCTAT GGTGATCCTT GCTCGGGGGT CTCCAGAGGA GGCTCATCGC TGGCCCCGTA 960  
TCACTCAGCC TGTCTTAA CGGCCTCGCC ATGTGCATTG TCACTTGTGC TGTCCAGATG 1020  
GGCACATGCA GCATGCTGTG CTCACAGCCC GCCGGCACGG CAGGTATGGG GGGTGTGACC 1080  
AAAATCAGTG GGATGTGGCA GGAAGCTGCA GCCCACGCCA GCATCTGTTT CCACAGGGAT 1140  
TTGTATCGTT GTGCCCCGTG CAGCTCCTGG GGAGATCTTT TACCTGTGCT TACTCCGTCT 1200  
GCGTTTCCTC CATCTACGGC TCAGGATCCC TCTGAGAGTT GATGAGGATG TGTAACAAGT 1260  
ATTTTCTTCT ATCGTGCCTG CCAGGGCTGA AGCTGCCTGG TATCCAGGAG GGAATGCTG 1320  
GTATCCCCAT ATGTCTGTGT TTGTTTGAGA TTTTAAATAA TAAATAATAA ATTTTGAAG 1380  
AATGAAAAA AAAAA 1395

(2) INFORMATION FOR SEQ ID NO: 106:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1635 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: PROSTUT10  
(B) CLONE: 1690990

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106 :

CCCTCTTCCT TTTGCGCAG GAAGAACAAA TCACAACAAT CACACACCAG GACTGAATCC 60



PF-0459 US

ATCAGCAGAT ACTGCCCTGT GGAAGGGCA GAGGAAAGAG AAGACAGACG GACTGACAGA 120  
CACCACAGAG GAACAGGGGA GTTAGCCTGG GACCAATGGA GGAGAAGTAC GAACCCTGGG 180  
AAAAAGACGT GTCAGATGAG AAAGTCCGG AGAGTCCGAT GTCTCATCGC AGGTGTTACA 240  
TCATCAGGGT TTGCCATTGG AATACTGAGT GGAGATGGGA AAGAGAAAAG TTAAGGGCTG 300  
AAATGGGAGG GGAATGGGAA GAAAAAATGA GAGACAAGAG GGAAATAAGA AAAAACAAAG 360  
AGAGCACAAA GACCAGTTTA GGAGAAAGGA CCAATGGGGA CAGTGGCAGA GTGGCGAGGT 420  
AGGTGAAGGA CTGAGGCACA GCGTCCTGTT GTGGAGGGAG GAAAGGCAAG CGTTCGAGG 480  
TGGTGAAAAG GAAGGCCTGC TAGGCACGGT GGGGATGAAC GAGGATGCCA TGAGTCACAC 540  
AAAAGACAGT GCTGGTGAGG CCCAGCCACA GGAGCCTCAG ATAAC TTGGT AAAGGCATGT 600  
CTCCCATTTG GGAAC TGATG TTCCTAAGAT CCGCACTGAC GCTGCTCAGC CGGTCCATCA 660  
CACAGCAAAG GCGTGAGGAA GGGTCACTGC CCAGCTGGAC TCCAGGGTGG TCCACGCATG 720  
ACAGTCACAC CGAACCTTCA TGAGGATGTG AACTGTTGGC TCCAATTTAC CATTCCCAGC 780  
AATTCCACTC AGATATTTGT ATACTAATGT TCACAGCAGC GTGAACTCCA CAGCAGGTGG 840  
AGTAATGTTC CATTGTGTGC ATATGCCACA TTTTGT TTAT CCATTCATCT GTTGATGCAC 900  
ATTTCGGTTG TTCCACCTT TGGGCTATTA TTAATAATGC TGCTGTGAAC ATTCCCAAGA 960  
GAAATAGGAA GACGGCTTTG CTAAGAACTA AAAAAGGGAT GGACAACAAG GGCATATACC 1020  
CAGGGGAGT GTTCTATCAT GACAGCTTTA CTGAGAGCAG AGTAGTTCTG CTCAGAATCA 1080  
GAACACTTGT TCCCTATAGC CCCCCTGATT GCCCCACAAC CACCACCGCA TACTCCCCTT 1140  
TTCCCAACCA TGGGCAGCAG ATTGAGCTAT TAACAGAAGT GTCCTTTTCGC TGGATTTCTC 1200  
AACCCTTTCC TCATCGTCCA CATAGAGAAA CAGTAACAGA TTGCTACTCA CCCAACACCC 1260  
AGGTCAAGTC CAATGCAGGT AGGAATAACA GCAAATCCTT CAATTTCTTG ATTCTGCTCT 1320  
TAAAAATCTT AACAGAGGCT TCCAGGTTCT GAAAATATTT TCTGCATAAA CGTGTGACAC 1380  
TCCATCACGA AACTCCCTTT GGTATFCTGC TTAACCTTAT CGCAAATGTC TGGAACGCTG 1440  
GTGGCTTCCA AAATCAACTC CTGGTGCTGC TTAATTAAGG TCAGGGCCAC CCGGAAGATA 1500  
ATCTTCGAGC CTTGTTAAA CAAACAGTCC CAGATCCGAA GCACTGTCTC CACGGGCAAG 1560  
ATGTCCACAA ACAGGCAGAT GAACCAGCGG GACACCAGCA GCGTCCACAG CACACCGAGA 1620  
CGCTCCATCA GGGGG 1635

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:

PF-0459 US

(A) LENGTH: 1485 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: DUODNOT02  
(B) CLONE: 1704050

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107 :

TTTTTGGTCC CGNCNAAAGN CCNAAAACCC GGNACCCGGG AAGCCNCCCC AANNCNAAAN 60  
TCCCAGTTN GAANCCCGAA GGNAAAACCC CGGAAAAGNA NNCNGCCCN AAANTTCNCG 120  
GGCNAAAACC CGGCCNTTTT TTCCCCCCCG GCGGCCCGTT TTGGGCCCCN GANTTTCCAT 180  
TTAAANTNCC NAGNCTTGGG CAACCTAACC AGGNTTTTCC CCCAANCTGG AAAAAGCCCG 240  
GCCAAGTTGA GCCGCACCCG CCCCAGAAGT TCAAGGGCCC CCGGCCTCCT GCGCTCCTGC 300  
CGCCGGGACC CTCGACCTCC TCAGAGCAGC CGGTGCGCGC CCCGGAAGA TGGCGAGGAG 360  
GAGCCGCCAC CGCCTCCTCC TGCTGCTGCT GCGCTACCTG GTGGTCGCCC TGGGCTATCA 420  
TAAGGCCTAT GGGTTTTCTG CCCCAAAAGA CCAACAAGTA GTCACAGCAG TAGAGTACCA 480  
AGAGGCTATT TTAGCCTGCA AAACCCCAA GAAGACTGTT TCCTCCAGAT TAGAGTGGAA 540  
GAAACTGGGT CGGAGTGTCT CCTTTGTCTA CTATCAACAG ACTCTTCAAG GTGATTTTAA 600  
AAATCGAGCT GAGATGATAG ATTTCAATAT CCGGATCAAA AATGTGACAA GAAGTGATGC 660  
GGGGAAATAT CGTTGTGAAG TTAGTGCCCC ATCTGAGCAA GGCCAAAACC TGGAAGAGGA 720  
TACAGTCACT CTGGAAGTAT TAGTGGCTCC AGCAGTTCCA TCATGTGAAG TACCCTCTTC 780  
TGCTCTGAGT GGAAGTGTGG TAGAGCTACG ATGTCAAGAC AAAGAAGGGA ATCCAGCTCC 840  
TGAATACACA TGGTTTAAGG ATGGCATCCG TTTGCTAGAA AATCCCAGAC TTGGCTCCCA 900  
AAGCACCAAC AGCTCATACA CAATGAATAC AAAAAGTGA ACTCTGCAAT TTAATACTGT 960  
TTCCAAACTG GACACTGGAG AATATTCCTG TGAAGCCCGC AATTCTGTTG GATATCGCAG 1020  
GTGTCCTGGG AAACGAATGC AAGTAGATGA TCTCAACATA AGTGGCATCA TAGCAGCCGT 1080  
AGTAGTTGTG GCCTTAGTGA TTTCCGTTTG TGGCCTTGGT GTATGCTATG CTCAGAGGAA 1140  
AGGCTACTTT TCAAAAGAAA CCTCCTTCCA GAAGAGTAAT TCTTCATCTA AAGCCACGAC 1200  
AATGAGTGAA AATGATTTC AAGCACACAAA ATCCTTTATA ATTTAAAGAC TCCACTTTAG 1260  
AGATACACCA AAGCCACCGT TGTTACACAA GTTATTAAAC TATTATAAAA CTCTGCTTTG 1320  
TCCGACATTT GCAAAGAGGT ACACGAGGAA ATGGAATTGG TATTTTATTT TAATTTTCAT 1380  
GACTACTAAC TCACCTGAAC TTGCTATTTT AAACAAATAG TTCTGTGAC ACCTAAAATA 1440  
TAATCTGGCT TCTTGTGTCT GGAATAAGTT AAAAGAATTA AAATA 1485

(2) INFORMATION FOR SEQ ID NO: 108:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 810 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: PROSNOT16  
 (B) CLONE: 1711840

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108 :

CGAGTGAGCG CGCGGCGGCC CCTGGTCCGC CCGGCCGCGG CCGATCTAGG GGCTGGGGGC 60  
 TGGAGGCGGG GGTGGGGGTC TGAGCTGCGT CCTGGGCTCG AGGCGTCCCC CGGGGAGTCG 120  
 CCTCTTAGCG GTGCGTCCGG GCTAGCGGCG AGGGGCCGCC CCAAGTCTTC CCACCGCCGC 180  
 CACCTTAGCA GCGCGACTTG GGGCCTGGAA AGTGGAGCAC GCGGAGGTGG GAGGGCCCTG 240  
 CACGCGGCCC CCGGTGGGGA AGGGGACGGG CCAGGGATTC AGACTCGGGC TCTCCCCTCA 300  
 GGATGCAGCA CCGAGGCTTC CTCCTCCTCA CCCTCCTCGC CCTGCTGGCG CTCACCTCCG 360  
 CGGTCGCCAA AAAGCAAGAT AAGGTGAAGA AGGGCGGCCC GGGGAGCGAG TGCCTGAGT 420  
 GGGCCTGGGG GCCCTGCACC CCCAGCAGCA AAGGATTTGC GGCAGTGGGT TTTCCGCGAG 480  
 GGCCACCTTG GGGGGGCCCA AGAACCCAAC CGGCAGTCCT GGTGAAAGG GTTGCCCTG 540  
 GAAAGTTGGA AAGAAAGGAG TTTTGGGCAC CCGGACTTTG GAAAGTTGGC CAAATTTTTT 600  
 GGAAGAAAAC TTGGCGGGTC TGCCGGTCCG TTAAATGGGG GAGGGGACAA AAGAATTGAA 660  
 AGCCGAAAAA ATGCTTTCTC CGCCGCCAAG AGAGGTCGAA CCCGCGTCTG GCAAGAAGAG 720  
 AAAAGGGCGC GCCCAGCTG TTAACAACAA TATGGCGCCT GAACAGTTGG TGGCACCACA 780  
 GGGGGAGGGA GACACATACT TGC GCGCGGT 810

(2) INFORMATION FOR SEQ ID NO: 109:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1064 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109 :

TTCCTGGGGC TCCGGGGCGC GGAGAAGCTG CATCCCAGAG GAGCGCGTCC AGGAGCGGAC 60

PF-0459 US

CCGGGAGTGT TTCAAGAGCC AGTGACAAGG ACCAGGGGCC CAAGTCCCAC CAGCCATGCA 120  
GACCTGCCCC CTGGCATTCC CTGGCCACGT TTCCCAGGCC CTTGGGACCC TCCTGTTTTT 180  
GGTGCCCTCC TTGAGTGCTC AGAATGAAGG CTGGGACAGC CCCATCTGCA CAGAGGGGGT 240  
AGTCTCTGTG TCTTGGGGCG AGAACACCGT CATGTCCTGC AACATCTCCA ACGCCTTCTC 300  
CCATGTCAAC ATCAAGCTGC GTGCCCACGG GCAGGAGAGC GCCATCTTCA ATGAGGTGGC 360  
TCCAGGCTAC TTCTCCCGGG ACGGCTGGCA GCTCCAGGTT CAGGGAGGCG TGGCACAGCT 420  
GGTGATCAAA GGCGCCCGGG ACTCCCATGC TGGGCTGTAC ATGTGGCACC TCGTGGGACA 480  
CCAGAGAAAT AACAGACAAG TCACGCTGGA GGTTCAGGT GCAGAACCCC AGTCCGCCCC 540  
CGACACTGGG TTCTGGCCTG TGCCAGCGGT GGTCACTGCT GTCTTCATCC TCTTGGTCGC 600  
TCTGGTCATG TTCGCCTGGT ACAGGTGCCG CTGTTCCCAG CAACGCCGGG AGAAGAAGTT 660  
CTTCCTCCTA GAACCCAGAG TGAAGGTCGC AGCCCTCAGA GCGGGAGCCC AGCAGGGCCT 720  
GAGCAGAGCC TCCGCTGAAC TGTGGACCCC AGACTCCGAG CCCACCCCAA GGCCGCTGGC 780  
ACTGGTGTTT AAACCCCTCAC CACTTGAGC CCTGGAGCTG CTGTCCCCC AACCCTTGTT 840  
TCCATATGCC GCAGACCCAT AGCCGCTGC AAGGAAGAGA GGACACAGGA GTAGCCACCC 900  
TGAGTGCCGA CCTTTGGTGG CGGGGGCCTG GGTCTCTCGT CCCCACCCGG AAGGGCACAA 960  
GACACCGGGC TTTGCTTGGC AAGGCTTGGG GCCTCTTGTC GTCAACCCAG TTCCCTTGGG 1020  
TGCCGTTGCA GAACCCCTTA GCCCCTTCCA ACGTCGACCA GGT 1064

(2) INFORMATION FOR SEQ ID NO: 110:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1031 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110 :

AGTTCCTGCA GGTGCCGGCG GTGACGCGGG CTTACACCGC AGCCTGTGTC CTCATCCACC 60  
GCCGCGGTGC AGCTGGAGCT CCTCAGCCCC TTTCAACTCT ACTTCAACCC GCACCTTGTG 120  
TTCCGGAAGT TCCAGGTGAG GCCGCCTCGC GCCGCGCACC TGGGGCCCGA CCCACCCACC 180  
CCGCACCTGA CCGCCCGTCC CCCGTAGGTC TGGAGGCTCG TCACCAACTT CCTCTTCTTC 240  
GGGCCCCTGG GATTCACTT CTTCTTCAAC ATGCTCTTCG TGTATCCTGC GCCTGCGGAC 300  
ACGGGCTGGG TGGAGGGCAG GCCGCGCGGG CTGGGAGAGA GGCCGGGACG GGGAAACTGA 360

PF-0459 US

GGCCCCGCCT GGTGGCACTT CCTATACCGA CGCCGTAGGT TCCGCTACTG CCGCATGCTG 420  
GAAGAGGGCT CCTTCCGCGG CCGCACGGCC GACTTCGTCT TCATGTTTCT CTTCGGGGGC 480  
GTCCTTATGA CCGTATCCTT CCCGCAGGCT CTGGAACCTC GGGCTAGGGC GCCTCGGCGT 540  
CCAGCCTGTG TTGGTCTCTG GGCCAACACA GCCATGCCAG AGAGGGACAC AGTCGCTGTC 600  
TCCAGCTTAG CACCGTTCCT GCCTTGGGCG CTCATGGGCT TCTCGCTGCT GCTGGGCAAC 660  
TCCATCCTCG TGGACCTGCT GGGGATTGCG GTGGGCCATA TCTACTACTT CCTGGAGGAC 720  
GTCTTCCCCA ACCAGCCTGG AGGCAAGAGG CTCCTGCAGA CCCCTGGCTT CCTAAAGCTG 780  
CTCCTGGATG CCCCTGCAGA AGACCCCAAT TACCTGCCCC TCCCTGAGGA ACAGCCAGGA 840  
CCCCATCTGC CACCCCCGCA GCAGTGACCC CCACCCAGGG CCAGGCCTAA GAGGCTTCTG 900  
GCAGCTTCCA TCCTACCCAT GACCCCTACT TGGGGCAGAA AAAACCCATC CTAAAGGCTG 960  
GGCCCATGCA AGGGCCCACC TGAATAAACA GAATGAGCTG CAAAAA AAAAAGGGC 1020  
GGCCGTCGCG A 1031

(2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2316 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PROSTUT12
- (B) CLONE: 1812375

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111 :

GCTGGATAAG ACACCAGGGG AGTCACTACA TGGTTACCGC ATCTGTATCC AGGCCATCCT 60  
GCAAGACAAG CCCAAGATTG CCACGGCAAA CCTAGGCAAG TTCCTGGAAC TGCTGAGGTC 120  
CCACCAGAGC CGACCAGCAA AGTGTCTCAC CATCATGTGG GCCCTGGGTC AAGCAGGTTT 180  
TGCCAACCTC ACCGAGGGAC TGAAAGTGTG GCTGGGGATC ATGCTGCCTG TGCTGGGCAT 240  
CAAGTCTCTG TCTCCCTTTG CCATCACATA CCTGGATCGG CTGCTCCTGA TGCATCCCAA 300  
CCTTACCAAG GGCTTCGGCA TGATTGGCCC CAAGGACTTC TTCCCACTTC TGGACTTTGC 360  
CTATATGCCG AACAACTCCC TGACACCCAG CCTGCAGGAG CAGCTGTGTC AGCTCTACCC 420  
CCGACTGAAA ATGCTGGCAT TTGGAGCAAA GCCGGATTCC ACCCTGCATA CCTACTTCCC 480  
TTCTTTCTCTG TCCAGAGCCA CCCCTAGCTG TCCCCCTGAG ATGAAGAAAG AGCTCCTGAG 540  
CAGCCTGACT GAGTGCCTGA CGGTGGACCC CCTCAGTGCC AGCGTCTGGA GGCAGCTGTA 600

PF-0459 US

CCCTAAGCAC CTGTCACAGT CCAGCCTTCT GCTGGAGCAC TTGCTCAGCT CCTGGGAGCA 660  
GATTCCCAAG AAGGTACAGA AGTCTTTGCA AGAAACCATT CAGTCCCTCA AGCTTACCAA 720  
CCAGGAGCTG CTGAGGAAGG GTAGCAGTAA CAACCAGGAT GTCGTACCT GTGACATGGC 780  
CTGCAAGGGC CTGTTGCAGC AGGTTTCAGG TCCTCGGCTG CCCTGGACGC GGCTCCTCCT 840  
GTTGCTGCTG GTCTTCGCTG TAGGCTTCCT GTGCCATGAC CTCCGGTCAC ACAGCTCCTT 900  
CCAGGCCTCC CTTACTGGCC GGTTCGCTCG ATCATCTGGC TTCTTACCTG CTAGCCAACA 960  
AGCGTGTGCC AAGCTCTACT CCTACAGTCT GCAAGGCTAC AGCTGGCTGG GGGAGACACT 1020  
GCCGCTCTGG GGCTCCCACC TGCTCACCGT GGTGCGGGCC AGCTTGCAGC TGGCCTGGGC 1080  
TCACACCAAT GCCACAGTCA GCTTCCTTTC TGCCCACTGT GCCTCTCACC TTGCGTGGTT 1140  
TGGTGACAGT CTCACCAGTC TCTCTCAGAG GCTACAGATC CAGCTCCCCG ATTCCGTGAA 1200  
TCAGCTACTC CGCTATCTGA GAGAGCTGCC CCTGCTTTTC CACCAGAATG TGCTGCTGCC 1260  
ACTGTGGCAC CTCTTGCTTG AGGCCCTGGC CTGGGCCCAG GAGCACTGCC ATGAGGCATG 1320  
CAGAGGTGAG GTGACCTGGG ACTGCATGAA GACACAGCTC AGTGAGGCTG TCCACTGGAC 1380  
CTGGCTTTGC CTACAGGACA TTACAGTGGC TTTCTTGGAC TGGGCACTTG CCCTGATATC 1440  
CCAGCAGTAG GCCCTGCCTT CCTGGCCACT GATTTCTGCA TGGGTAGACC ATCCAAGACT 1500  
GCAGCGGGTA GAAGGTGGCA GTTCTTCATG GGAGTCTTTT TAACTTGGTG CCTGAGTTCT 1560  
CTCCTAGGCA AGTGGCCAGT TGCCTCCACC TCAGTTCTTC CATCTTTGGT GGGGACAGGG 1620  
CCCAGCAGCA TCTCAGCCTC CTACCCACAA TTCCACTGAA CACTTTTCTG GCCCTACTGC 1680  
ACATGGCCCC CAGCCTCCAT CCTTGTGCTG GTAGCCTCTC ACAACTCCGC CCTTGCCCTC 1740  
TGCCTTCCAC TTCCTTCCAT CTCATTTCTA AACCCCAAAC AGCTCATCTC TAAAAAGATA 1800  
GAACTCCCAG CAGGTGGCTT CTGTGTTCTT CTGACAAATG ATTCCTGCTT CTCCAGACTT 1860  
TAGCAGCCTC CTGTTCCCAT TCTTGCTCAC AGCTCTAGCC ACAGCAGAAG GAAAGGGGCT 1920  
TCCAGAAGAA TATAGCACCG CATTGGGAAA CAGCAGCCTC ACCTCCACCT GAAGCCTGGG 1980  
TGTGGCTGTC AGTGGACATG GGGAGCTGGA TGGAAATGCC TCTCACTTCA AAATGCCCAG 2040  
CCTGCCCAA ATGCCTCTAA GCCCCTCCCT GTCCCCTCCC TTGTAGTCCT ACTTCTTCCA 2100  
ACTTTCCATT CCCCATCATG CTGGGGGTCT TGGTCACAAG GCTCAGCTTC TCTCCACTGT 2160  
CCATCCCTCC TATCATCTGT AGAGCAGAGC ACAGGCAGTT GTGTGCCTTG GGCCAGGGA 2220  
ACCCTCCATC AACCTGAGAC AGGACTCAGT ATATGTTTCT TGGGTATGCC CTACCAGGTG 2280  
GAATAAAGGA CACAGATTTG AAAAAAAAAA AAAAAA 2316

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1169 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: PROSNOT20  
 (B) CLONE: 1818761

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112 :

```

AGCAAGGAGC CAGAGGCCAT GCAGTGGCTC AGGGTCCGTG AGTCGCCTGG GGAGGCCACA 60
GGACACAGGG TCACCATGGG GACAGCCGCC CTGGGTCCCG TCTGGGCAGC GTCCTGCTC 120
TTTCTCCTGA TGTGTGAGAT CCCTATGGTG GAGCTCACCT TTGACAGAGC TGTGGCCAGC 180
GGCTGCCAAC GGTGCTGTGA CTCTGAGGAC CCCCTGGATC CTGCCCATGT ATCCTCAGCC 240
TCTTCCTCCG GCCGCCCCCA CGCCCTGCCT GAGATCAGAC CCTACATTAA TATCACCATC 300
CTGAAGGGTG ACAAAGGGGA CCCAGGCCCA ATGGGCCTGC CAGGGTACAT GGGCAGGGAG 360
GGTCCCCAAG GGGAGCCTGG CCCTCAGGGC AGCAAGGGTG ACAAGGGGGA GATGGGCAGC 420
CCCGGCGCCC CGTGCCAGAA GCGCTTCTTC GCCTTCTCAG TGGGCCGCAA GACGGCCCTG 480
CACAGCGGCG AGGACTTCCA GACGCTGCTC TTCGAAAGGG TCTTTGTGAA CCTTGATGGG 540
TGCTTTGACA TGGCGACCGG CCAGTTTGCT GCTCCCCTGC GTGGCATCTA CTCTTCAGC 600
CTCAATGTGC ACAGCTGGAA TTACAAGGAG ACGTACGTGC ACATTATGCA TAACCAGAAA 660
GAGGCTGTCA TCCTGTACGC GCAGCCCAGC GAGCGCAGCA TCATGCAGAG CCAGAGTGTG 720
ATGCTGGACC TGGCCTACGG GGACCGCGTC TGGGTGCGGC TCTTCAAGCG CCAGCGCGAG 780
AACGCCATCT ACAGCAACGA CTTCGACACC TACATCACCT TCAGCGGCCA CCTCATCAAG 840
GCCGAGGACG ACTGAGGGCC TCTGGGCCAC CCTCCCGGCT GGAGAGCTCA GGTGCTGGTC 900
CCGTCCCCTG CAGGGCTCAG TTTGCACTGC TGTGAAGCAG GAAGGCCAGG GAGGTCCCCG 960
GGGACCTGGC ATTCTGGGGA GACCCTGCTT CTATCTTGGC TGCCATCATC CCTCCCAGCC 1020
TATTTCTGCT CCTCTCTTCT CTCTTGGACC TATTTTAAGA AGCTTGCTAA CCTAAATATT 1080
CTAGAACTTT CCCAGCCTCG TAGCCCAGCA CTTCTCAAAC TTGGAAATGC ATGCGAATCA 1140
CCCGGGGTTC GTGTTAAATG CAGATTCTG                                     1169
  
```

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1530 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: GBLATUT01  
(B) CLONE: 1824469

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113 :

TCACAGACTG CGGAGTGGGT CAGGGGCTGC GAGGGCTGCC CCAAGTCCTA CCGGGTTTGC 60  
ACGGGCGCGC CCGGCTCCGC CCGCAAGTGC GCCTTCCTGA CTTACTGCTG GGTGCGCGGG 120  
GCTGGGGGTG CGAGTACCAC CCCTGAAGTC TCTTCCTGGG CGACCTCCGG GGCCTCATTC 180  
TAGGCCTCCT TAAAGAGAAG GATCTAAATT AGGAAAAGGA AGTGCCCTTA TCCACGACCA 240  
AGCTCTTCCA CCTGCGGAGC TCGCTTAGTC TGCACCTCAA CCGTGCGGAA AGTGA CTGCC 300  
CTGTTTACTG AGGAAAAACT GGGGCTCAGA AAGATACCAT GAGTAGTTTG AAACAGGAAC 360  
AAAATCTTCT GAAAGCTCGG AGCAGAAGCC TTTTGGTCA ACATGGAGGA AAAAAGACGG 420  
CGAGCCCGAG TTCAGGGAGC CTGGGCTGCC CCTGTAAAAA GCCAGGCCAT TGCTCAGCCA 480  
GCTACCACTG CTAAGAGCCA TCTCCACCAG AAGCCTGGCC AGACCTGGAA GAACAAAGAG 540  
CATCATCTCT CTGACAGAGA GTTTGTGTTC AAAGAACCTC AGCAGGTAGT ACGTAGAGCT 600  
CCTGAGCCAC GAGTGATTGA CAGAGAGGGT GTGTATGAAA TCAGCCTGTC ACCCACAGGT 660  
GTATCTAGGG TCTGTTTGTA TCCTGGCTTT GTTGACGTGA AAGAAGCTGA CTGGATATTG 720  
GAACAGCTTT GTCAAGATGT TCCCTGGAAA CAGAGGACCG GCATCAGAGA GGATATAACT 780  
TATCAACAAC CAAGACTTAC AGCATGGTAT GGAGAACTTC CTTACACTTA TTCAAGAATC 840  
ACTATGGAAC CAAATCCTCA CTGGCACCTT GTGCTGCGCA CACTAAAGAA CCGCATTGAA 900  
GAGAACACTG GCCACACCTT CAACTCCTTA CTCTGCAATC TTTATCGCAA TGAGAAGGAC 960  
AGCGTGGACT GGCACAGTGA TGATGAACCC TCACTAGGGA GGTGCCCCAT TATTGCTTCA 1020  
CTAAGTTTTG GTGCCACACG CACATTTGAG ATGAGAAAGA AGCCACCACC AGAAGAGAAT 1080  
GGAGACTACA CATATGTGGA AAGAGTGAAG ATACCCTTGG ATCATGGTAC CTTGTTAATC 1140  
ATGGAAGGAG CGACACAAGC TGA CTGGCAG CATCGAGTGC CCAAAGAATA CCACTCTAGA 1200  
GAACCGAGAG TGAACCTGAC CTTTCGGACA GTCTATCCAG ACCCTCGAGG GGCACCCTGG 1260  
TGACGTCAGA GCTTTGAGAG AGAAGCTTCA CTGAAACGGA GCAAACCTTC CACTGAGAAG 1320  
CCACTTCAAG AGGCTGGTGC TGCTAGATCT CATGATGTGG CTGTTGGGAA GATGGTGGGG 1380  
TTTGTGTTGCC AGCTTGAGT CCTATTAAAT GAAAGCCAGC AACTCATGTT GGTAATAGGT 1440  
CTACTGTGGG AACAGTTATC CCTAACCACA GCTCAAATC GCTATCATCT TTAGGCAAAT 1500



TAAATCTAT GTGGCAGTGA AAAAAAAAAA

1530

## (2) INFORMATION FOR SEQ ID NO: 114:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1336 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PROSNOT19
- (B) CLONE: 1864292

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114 :

AGCTCGTACC CCTCGAGTGA AATTCTGAAA TGAAGATGGA GGAGGCAGTG GGAAAAGTTG 60  
AAGAACTCAT TGAGTCCGAA GCCCCACCAA AAGCATCTGA ACAAGAGACA GCCAAGGAGG 120  
AAGATGGATC TGTAGAACTG GAATCTCAAG TTCAGAAAGA TGGTGTAGCG GATTCTACAG 180  
TTATTTCTTC AATGCCCTGC TTGTTGATGG AACTGAGAAG GGACTCTTCT GAGTCTCAGT 240  
TAGCATCCAC AGAGAGTGAC AAGCCTACAA CTGGCCGAGT TTATGAGAGT GACCCCTCTA 300  
ATCACTGCAT GCTTTCCCCT TCCTCTAGTG GTCACCTGGC TGATTCAGAT ACGTTGTCTT 360  
CCGCAGAAGA GAATGAACCC TCTCAGGCAG AAACGGCGGT AGAAGGAGAC CCTTCAGGAG 420  
TGTCTGGTGC CACAGTTGGG CGCAAGTCTA GGCGGTCCCG ATCTGAAAGT GAAACTTCCA 480  
CTATGGCTGC CAAGAAAAAC CGGCAATCCA GTGATAAACA GAATGGCCGA GTCGCCAAGG 540  
TTAAAGGTCA TCGGAGCCAA AAGCACAAGG AGAGGATCAG GCTACTGAGG CAGAAACGGG 600  
AGGCTGCTGC AAGGAAGAAA TATAACCTGC TGCAGGACAG TAGTACCAGT GATAGTGACC 660  
TGACTTGTGA CTCAAGCACG AGCTCATCAG ATGATGATGA AGAGGTTTCA GGGAGCAGCA 720  
AGACAATCAC TGCAGAGATA CCAGATGGAC CTCCAGTTGT AGCTCATTAT GATATGTCTG 780  
ACACCAACTC TGACCCAGAA GTGGTAAATG TGGACAATTT ATTGGCGGCT GCAGTAGTTC 840  
AAGAGCACAG TAATTCTGTA GGCGGCCAGG ACACAGGAGC TACCTGGAGG ACCAGCGGGC 900  
TTCTAGAGGA GCTGAATGCA GAGGCAGGTC ATTTGGATCC AGGATTCCTA GCAAGTGACA 960  
AAACATCTGC TGGCAATGCG CCACTCAATG AAGAAATTAA CATTGCGTCT TCAGATAGTG 1020  
AAGTAGAGAT TGTGGGAGTT CAGGAACATG CAAGGTGTGT TCATCCTCGA GGTGGTGTGA 1080  
TTCAGAGTGT TTCTTCATGG AAGCATGGCT CGGGCACGCA GTATGTTAGC ACCAGGCAAA 1140  
CACAGTCATG GACTGCTGTG ACTCCCCAGC AGACTTGGGC TTCACCAGCA GAAGTTGTTG 1200  
ACCTTACCTT GGATGAGGAT AGCAGGCGTA AATACCTACT GTAATACAAT GTCAGTGTGT 1260

PF-0459 US

TTCCTCTGCA CTGTTCCCTT CCACTTCCTC ATCCTCTTTG TGACATGGAA GTTCATTGTC 1320  
ATAGGGGTAC GGAGCT 1336

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1742 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: THP1NOT01
- (B) CLONE: 1866437

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115 :

CCCCCGCCCC CTCCCCGCCC GCCTTCCCGG TGACCTTCAG GGGCCCGGGT GGCGGGCGCA 60  
GGCCCCTGCG GCGGCGGCGG GATGTTCTGT CAGGAGGAGA AGATCTTCGC GGGCAAGGTG 120  
CTGCGGCTGC ACATCTGCGC GTCCGACGGC GCCGAGTGGC TGGAGGAGGC CACCGAGGAC 180  
ACCTCGGTGG AGAAGCTCAA GGAGCGCTGC CTCAAGCACT GTGCTCATGG GAGCTTAGAA 240  
GATCCCAAAA GTATAACCCA TCATAAATTA ATCCACGCTG CCTCAGAGAG GGTGCTGAGT 300  
GATGCCAGGA CCATCTGGA AGAGAACATC CAGGACCAAG ATGTCCTATT ATTGAAAAAA 360  
AAGCGTGCTC CATCACCCT TCCCAAGATG GCTGATGTCT CAGCAGAAGA AAAGAAAAAA 420  
CAAGACCAGA AAGCTCCAGA TAAAGAGGCC ATACTGCGGG CCACCGCCAA CCTGCCCTCC 480  
TACAACATGG ACCGGGCCGC GGTCCAGACC AACATGAGAG ACTTCCAGAC AGAACTCCGG 540  
AAGATACTGG TGTCTCTCAT CGAGGTGGCG CAGAAGCTGT TAGCGCTGAA CCCAGATGCG 600  
GTGGAATTGT TTAAGAAGGC GAATGCAATG CTGGACGAGG ACGAGGATGA GCGTGTGGAC 660  
GAGGCTGCCC TCGGCGAGCT CACGGAGATG GGCTTTCCGG AGAACAGAGC CACCAAGGCC 720  
CTTCAGCTGA ACCACATGTC GGTGCCTCAG GCCATGGAGT GGCTAATTGA ACACGCAGAA 780  
GACCCGACCA TAGACACGCC TCTTCCTGGC CAAGCTCCCC CAGAGGCCGA GGGGGCCACA 840  
GCAGCTGCCT CCGAGGCTGC CGCGGGAGCC AGCGCCACCG ATGAGGAGGC CAGAGATGAG 900  
CTGACGGAAA TCTTCAAGAA GATCCGGAGG AAAAGGGAGT TTCGGGCTGA TGCTCGGGCC 960  
GTCATTTCCC TGATGGAGAT GGGGTTCGAC GAGAAAGAGG TGATAGATGC CCTCAGAGTG 1020  
AACAACAACC AGCAGAATGC CGCGTGCGAG TGGCTGCTGG GGGACCGGAA GCCCTCTCCG 1080  
GAGGAGCTGG ACAAGGGCAT CGACCCGAC AGTCCTCTCT TTCAGGCCAT CCTGGATAAC 1140  
CCGGTGGTGC AGCTGGGCCT GACCAACCCG AAAACATTGC TAGCATTTGA AGACATGCTG 1200

PF-0459 US

GAGAACCCAC TGAACAGCAC CCAGTGGATG AATGATCCAG AACGCGGCC TGTCATGCTG 1260  
CAGATCTCTA GAATCTTCCA GACACTAAAT CGCACGTAGG TGGCGTTGTT CCACTCGGCT 1320  
ATCAGGCCAC AGCAGCCCCC TGGTGC GGCC CGAGACCGGG CAGAGTGGAC CTCACCTGGA 1380  
AACTCACCTT CAGCGCCTCA GCCCTGGACT GTTAGAGGTG CTGCAGCTGC TCCTGCTCTC 1440  
TGATCTTATT GCTTATAAAC TTTGGTGACG GTAGTGTGTA AGGCCGTATT TTTAGCATCT 1500  
GACAGGTGTT TACAAAAAAG TGGTTGTCGC ACTGGGAAGT GGAGTGATGG CCTCGTCTCC 1560  
AGTGCTCCTC TGGGCTCTTG AGTTGCTGCT TGAATTGCCG TGTAGACATT TGCTTGAGAGA 1620  
GTCCACTTGT TATTTGACGG AGGTAGGTTT CAACCCAGAG TTAATGTCAA GCATGCTAAT 1680  
TTAACTAGTC ACTCACAGAT GACTTTTCTT TAATAAAGTC CCTTTTCCTA TTAACAAAAA 1740  
AA 1742

(2) INFORMATION FOR SEQ ID NO: 116:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1074 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: SKINBIT01  
(B) CLONE: 1871375

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116 :

GCGGTGCAGA GGAAGCACAA CCTCTACCGG GACAGCATGG TCATGCACAA CAGCGACCCC 60  
AACCTGCACC TGCTGGCCGA GGGCGCCCCC ATCGACTGGG GCGAGGAGTA CAGCAACAGC 120  
GGCGGGGGCG GCAGCCCAGC CCCAGCACCC CGGAGTCAGC CACCCTCTCG GAAAAGCGAC 180  
GGCGCGCCAA GCAGGTGGTC TCTGTGGTCC AGGATGAGGA GGTGGGGCTG CCCTTTGAGG 240  
CTAGCCCTGA GTCACCACCA CCTGCGTCCC CGGACGGTGT CACTGAGATC CGAGGCCTGC 300  
TGGCCCAAGG TCTGCGGCCT GAGAGCCCCC CACCAGCCGG CCCCCTGCTC AACGGGGCCC 360  
CCGCTGGGGA GAGTCCCCAG CCTAAGGCCG CCCCCGAGGC CTCCTCGCCG CCTGCCTCAC 420  
CCCTCCAGCA TCTCCTGCCT GGAAAGGCTG TGGACCTTGG GCCCCCAAG CCCAGCGACC 480  
AGGAGACTGG AGAGCAGGTG TCCAGCCCCA GCAGCCACCC CGCCCTCCAC ACCACCACCG 540  
AGGACNANTT TCAAGGGGTG CAAGAATTGA AGNTTCNTAA GGGCCAANTT GGGGGTCCCC 600  
TTGACTTGGN TTGNAANAT TGGGGCAAAA AGGGCCGGTT TTCCCNTTT CCCGGGANAC 660  
CCCAAGGGAA AGGGGNTTCA AAGCTTCTTN GGGGGGGAAG GGGGGAANCC CTTGGGTNTT 720

PF-0459 US

TTGTTGGCCN TTTGTGANCA NCAGCGAGGA GAGTGCAAAG GTGCAGAGTN AGTTNTAGGN 780  
CANTGGGTCC CTGACTGCTG CANATGGTAA GGNCGTTNNC TTGTGGACCC AAGGCAGGNA 840  
AAGNTGTGGG GAGGGAAGCT GGTNTGTGCN TTGTGGGTGG AAGCGGGGAN GGCTGTGTTG 900  
NANGGCAGGG AGAGGGCNAA NTGAGTTATT TATTGGGGTT CANGTGAAAA GTTTCCTGNN 960  
CCCTGTNTTG TGTNCTGTG GGATTGATTN TAAGATNGNN AGGGGTNGGT TTTTGGGGTT 1020  
TTCCTGGTTG GTGGCCAAAN GGTTTGAAA ATNGNTGGGG GGGNNTGGA NAAT 1074

(2) INFORMATION FOR SEQ ID NO: 117:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1454 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: LEUKNOT03  
(B) CLONE: 1880830

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117 :

CCCCGGGGGAG GCCTGACCCC CTCCGCACCA CCGTACGGAG CCGCATTTC CCGTTTCCC 60  
GAGGGGCATC CAGCCGTGTT GCCTGGGGAG GACCCACCCC CCTATTACC CTTAACTAGC 120  
CCGGACAGTG GGAGTGCCCC TATGATCACC TGCCGAGTCT GCCAATCTCT CATCAACGTG 180  
GAAGGCAAGA TGCATCAGCA TGTAGTCAA TGTGGTGTCT GCAATGAAGC CACCCCAATC 240  
AAGAATGCAC CCCCAGGGAA AAAATATGTT CGATGCCCT GTAACTGTCT CTTATCTGC 300  
AAAGTGACAT CCCAACGGAT TGCATGCCCT CGGCCCTACT GCAAAGAAT CATCAACCTG 360  
GGGCCTGTGC ATCCCGGACC TCTGAGTCCA GAACCCCAAC CCATGGGTGT CAGGGTTATC 420  
TGTGGACATT GCAAGAATAC TTTTCTGTGG ACAGAGTTCA CAGACGCAC TTTGGCACGT 480  
TGTCCTCACT GCAGGAAAGT GTCATCTATT GGGCGCAGAT ACCCACGTAA GAGATGTATC 540  
TGCTGCTTCT TGCTTGGCTT GCTTTTGGCA GTCAGTCCA CTGGCCTTGC CTTTGGCACA 600  
TGGAAGCATG CACGGCGATA TGGAGGCATC TATGCAGCT GGGCATTGT CATCCTGTTG 660  
GCTGTGCTGT GTTTGGGCCG GGCTCTTTAT TGGGCCTGTA TGAAGGTCAG CCACCCTGTC 720  
CAGAACTTCT CCTGAGCCTG ATGACCCACA GACTGTGCCT GGCCCTCCC TGGTGGGGAC 780  
AGTGACACTA CGAAGGGAGC TGGGCTAGTT AAAGGCTCCC GGGGCTTCTA GAAGGAAGCC 840  
AAGCAGCTGC CTTCTTTTC CCTGGGGAGA GGTAGGAAGG AACCAGGCC TCACTTAGGT 900  
TTGGAGGGGC AGATAAGAGC ACTGCTGACC ATCTGCTTTC CTCCAAGGGT TGCTGTGTCT 960

PF-0459 US

AGGGTGAAGT AGGCAAAACG TTGCCCTTAA AACTGGGCCC TGAAGACGGT TCCAGCCTTG 1020  
TCCTTCCTGT GTGCTCCCTG AGAGCCATTC CTGTCCCTTA CACATTCCAG GGCAGGGTGG 1080  
GGGTGGGTAG CCCTGGGGGT TCCCCTCCCT CTTGTGCACC ATTAGGACTT TGCTGCTGCT 1140  
ATTGCACTTC ACCAGAGGTT GGCTCTGGCC TCAGTACCCT CAGTCTCCTC TCCCCACATT 1200  
GTGTCCTGTG GGGGTGGGGT CAGCCGCTGC TCTGTACAGA ACCACAGGAA CTGATGTGTA 1260  
TATAACTATT TAATGTGGGA TATGTTCCCC TATTCCTGTA TTTCCCTTAA TTCCTCCTCC 1320  
CGACCTTTTT TACCCCCCA GTTGCAGTAT TTAAGTGGGC TGGGTAGGGT TGCTCAGTCT 1380  
TTGGGGGAGG TTAGGGACTT ATCCTGTGCT TGTAATAAAA TAAGGTCATG ACTCTAAAAA 1440  
AAAAAAAAGG GCGG 1454

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2071 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: OVARNOT07
- (B) CLONE: 1905325

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118 :

AGCTTTGAAT TCCTGTATCT GAGAACGGAT CGTTCGAGGT GGTGGAGGGG GTTGAATTG 60  
GGGACCTACG GAAGGCTCAG CTCTTGCCAG GCCAAATTGA GACATGTCTG ACACAAGCGA 120  
GAGTGGTGCA GGTCTAACTC GCTTCCAGGC TGAAGCTTCA GAAAAGGACA GTAGCTCGAT 180  
GATGCAGACT CTGTTGACAG TGACCCAGAA TGTGGAGGTC CCAGAGACAC CGAAGGCCTC 240  
AAAGGCACTG GAGGTCTCAG AGGATGTGAA GGTCTCAAAA GCCTCTGGGG TCTCAAAGGC 300  
CACAGAGGTC TCAAAGACCC CAGAGGCTCG GGAGGCACCT GCCACCCAGG CCTCGTCTAC 360  
TACTCAGCTG ACTGATACCC AGGTTCTGGC AGCTGAAAAC AAGAGTCTAG CAGCTGACAC 420  
CAAGAAACAG AATGCTGACC CGCAGGCTGT GACAATGCCT GCCACTGAGA CCAAAAAGGT 480  
CAGCCATGTG GCTGATACAA AGGTCAATAC AAAGGCTCAG GAGACTGAGG CTGCACCCTC 540  
TCAGGCCCCA GCAGATGAAC CTGAGCCTGA GAGTGCAGCT GCCCAGTCTC AGGAGAATCA 600  
GGATACTCGG CCCAAGGTCA AAGCCAAGAA AGCCCGAAAG GTGAAGCATC TGGATGGGGA 660  
AGAGGATGGC AGCAGTGATC AGAGTCAGGC TTCTGGAACC ACAGGTGGCC GAAGGGTCTC 720  
AAAGGCTCTA ATGGCCTCAA TGGCCCGCAG GTTTCAAGGG GTCCCATAGC CTTTTGGGCC 780

PF-0459 US

CGCAGGATTC AAGGACTCGG TTGGCTGCTT GGGCCCGGAG AGCCTTGCTC TCCCTGAGAT 840  
CACCTAAAGC CCGTAGGGCA AGGCTCGCCG TAGAGCTGCC AAGCTCCAGT CATCCCAAGA 900  
GCCTGAAGCA CCACCACCTC GGGATGTGGC CCTTTTGCAA GGGAGGGCAA ATGATTGGT 960  
GAAGTACCTT TTGGCTAAAG ACCAGACGAA GATTCCCATC AAGCGCTCGG ACATGCTGAA 1020  
GGACATCATC AAAGAATACA CTGATGTGTA CCCCCAAATC ATTGAACGAG CAGGCTATTC 1080  
CTTGAGAAAG GTATTTGGGA TTCAATTGAA GGAAATTGAT AAGAATGACC ACTTGACAT 1140  
TCTTCTCAGC ACCTTAGAGC CCACTGATGC AGGCATACTG GGAACGACTA AGGACTCACC 1200  
CAAGCTGGGT CTGCTCATGG TGCTTCTTAG CATCATCTTC ATGAATGGAA ATCGGTCCAG 1260  
TGAGGCTGTC ATCTGGGAGG TGCTGCGCAA GTTGGGGCTG CGCCCTGGGA TACATCATTC 1320  
ACTCTTTGGG GACGTGAAGA AGCTCATCAC TGATGAGTTT GTGAAGCAGA AGTACCTGGA 1380  
CTATGCCAGA GTCCCCAATA GCAATCCCCC TGAATATGAG TTCTTCTGGG GCCTGCGCTC 1440  
TTACTATGAG ACCAGCAAGA TGAAAGTCCT CAAGTTTGCC TGCAAGGTAC AAAAGAAGGA 1500  
TCCCAAGGAA TGGGCAGCTC AGTACCGAGA GGCGATGGAA GCAGATTTGA AGGCTGCAGC 1560  
TGAGGCTGCA GCTGAAGCCA AGGCTAGGGC CGAGATTAGA GCTCGAATGG GCATTGGGCT 1620  
CGGCTCGGAG AATGCTGCCG GGCCCTGCAA CTGGGACGAA GCTGATATCG GACCCTGGGC 1680  
CAAAGCCCGG ATCCAGGCGG GAGCAGAAGC TAAAGCCAAA GCCCAAGAGA GTGGCAGTGC 1740  
CAGCACTGGT GCCAGTACCA GTACCAATAA CAGTGCCAGT GCCAGTGCCA GCACCACTGG 1800  
TGGCTTCAGT GCTGGTGCCA GCCTGACCGC CACTCTCACA TTTGGGCTCT TCGCTGGCCT 1860  
TGGTGGAGCT GGTGCCAGCA CCAGTGGCAG CTCTGGTGCC TGTGGTTTCT CCTACAAGTG 1920  
AGATTTTAGA TATTGTAAAT CCTGCCAGTC TTTCTCTTCA AGCCAGGGTG CATCCTCAGA 1980  
AACCTACTCA ACACAGCACT CTAGGCAGCC ACTATCAATC AATTGAAGTT GACACTCTGC 2040  
ATTAAATCTA TTTGCCATTT CAAAAAAAAA A 2071

(2) INFORMATION FOR SEQ ID NO: 119:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1236 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: BRSTTUT01  
(B) CLONE: 1919931

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119 :

PF-0459 US

ACCTGGGACC CCCAGAACGG CCGCCCCTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 60  
TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTAG AAGGTTGAAA CCAGGCTTAT 120  
TTATTTTCAT CTTCTTCTG CCATCTTTTA ACCAACCTTC TCAGAATAAA ATGTGATTTT 180  
TGAGACAGAA TGAAACACAT ATCCAAATTT TAATACAGTA AGAATAGGTA TCCTGAATAA 240  
ATGAGAACTC TAGAAAATCA AGGTTTCAAA ATTCTACCTC TCCTGGGAGT TAAAGAAGTT 300  
TGGCAGAAAC AGAACAAATT AATCAGCAGA TTCATCACCT GCCAATTTTT TCTGTACAAT 360  
TTTCTTGATT CTGGGAGCAT CTGGGTCCAG GCAGATTTTC CTCCCATCCT TCAGTGTGGC 420  
TGCTTCTTGT TTCATCCATG GACCCTGCAA GAAATTGCCC CATGTTTCTG TTTGTGCATC 480  
ACTGAGAAAG GAAGCATGAA GGTGCGACAG GTCAGGCCAT TCCATTGCCC TCCTGGTGCC 540  
GGGTTTGCCC TCCCAATCCT GGGGTGCTT CAGGGGCTTG TCATTCTCCA TAGTCCCCTC 600  
CACATTTCTC AGGTTTCTGC TCAAAGTCA CCTTTTGGAG GGGTCTCCAC CTGTCACTGT 660  
GTTTGTAAGA GTCCTTCAG TTTCTTTCTA GTCATCTCA CTCTGGTAAT GTCTTTGATT 720  
ACCACCACCA TCTGACCTGG TCTTATGACC TGTTAGCTTT CTCATCAGA CGTGAGCACC 780  
AGGATGGCAG GGGCCTCATC TGTCTGTTC CTCCTGTGGC CTGGGTCCTA GCACCATGTC 840  
TGGTACAGTG TAGATGCTCA AGGGAAGTTT ACTTTGTAAA ACCACTTACC TGGGAGATGT 900  
TACTGTTAGT CTAACCTGTA CCATTTTGTA AACCTCCAGC CATTTTGCAG ACTCTGATCA 960  
CAGTGAAACG TTCCATGGGA ACTTGGGCCA TGAGAAACAT CCTTCCTAAC CACGTGACTG 1020  
CAGAAACATC CTTATCGCGT CCTCCTGGGC AAAGGCCCAA CAGCCTGACT GCAGGGACAT 1080  
CCTTGCCATA TCCTGCTGGG CAGCAAGCTC TACCACCCAG ATCCCTCCCT CCCAGTCCCA 1140  
TGATTACCCC AGCCTGTGAG TGGCAGTTGG TGCTGGCACT AAGCTGGTTT CCTCCTCCCC 1200  
AGGGTTTTGC TGGCAATAAA GATGTTGCTG TTGAAG 1236

(2) INFORMATION FOR SEQ ID NO: 120:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1391 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: BRSTNOT04  
(B) CLONE: 1969426

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120 :

GTACTGCCCCA CCACCTCCCT GGGCCACCCC TCACTCAGTG CTCCGGCTCT CTCCTCCTCC 60

PF-0459 US

TCTTCGTCCT CCTCCACTTC ATCTCCTGTT TTGGGCTCCC CCTCTTACCC TGCTTCTTCC 120  
CCTGGGGCCT CCCCCACCA CCGCCGTGTG CCCCTCAGCC CCCTGAGTTT GCTCGCGGGC 180  
CCAGCCGACG CCAGAAGGTC CCAACAGCAG CTGCCCCAAC AGTTTTTCGCC AACAAATGTCA 240  
CCCACCTTGT CTTCCATCAC TCAGGGCGTC CCCCTGGATA CCAGTAAACT GTCCACTGAC 300  
CAGCGGTTAC CCCCATACCC ATACAGCTCC CCAAGTCTGG TTCTGCCTAC CCAGCCCCAC 360  
ACCCCAAAGT CTCTACAGCA GCCAGGGCTG CCCTCTCAGT CTTGTTTCACT GCAGTCCTCA 420  
GGTGGGCAGC CCCCAGGCAG GCAGTCTCAT TATGGGACAC CGTACCCACC TGGGCCCAGT 480  
GGGCATGGGC AACAGTCTTA CCACCGGCCA ATGAGTGACT TCAACCTGGG GAATCTGGAG 540  
CAGTTCAGCA TGGAGAGCCC ATCAGCCAGC CTGGTGCTGG ATCCCCCTGG CTTTTCTGAA 600  
GGGCCTGGAT TTTTAGGGGG TGAGGGGCCA ATGGGTGGCC CCCAGGATCC CCACACCTTC 660  
AACCACCAGA ACTTGACCCA CTGTTCCCGC CATGGCTCAG GGCCTAACAT CATCCTCACA 720  
GGGGACTCCT CTCCAGGTTT CTCTAAGGAG ATTGCAGCAG CCCTGGCCGG AGTGCCTGGC 780  
TTTGAGGTGT CAGCAGCTGG ATTGGAGCTA GGGCTTGGGC TAGAAGATGA GCTGCGCATG 840  
GAGCCACTGG GCCTGGAAGG GCTAAACATG CTGAGTGACC CCTGTGCCCT GCTGCCTGAT 900  
CCTGCTGTGG AGGAGTCATT CCGCAGTGAC CGGCTCCAAT GAGGGCACCT CATCACCATC 960  
CCTCTTCTTG GCCCCATCCC CCACCACCAT TCCTTTCTCTC CCTTCCCCCT GGCAGGTAGA 1020  
GACTCTACTC TCTGTCCCCA GATCCTCTTT CTAGCATGAA TGAAGGATGC CAAGAATGAG 1080  
AAAAAGCAAG GGGTTTGTCC AGGTGGCCCC TGAATTCTGC GCAAGGGATG GGCCTGGGGG 1140  
AACTCAAGGG AGGGCCTAAA GCACTTGTA CTTTGAACCG TCTGTCTGGA GGTGAGAGCC 1200  
TGTTGGAAG CAGGGGTAGA GGGGAGCCCT GGAAGCAGGG CTTTTCCGGA TGCCTAGGGG 1260  
TGGGCAGTGC CAGCCCCTCC TCACCACTCT TCCCCTTGCA GTGGAGGAGA GAGCCAGAGT 1320  
GGATACTATT TTTTATTAAA TATATTATTA TATGTTAATA AAAAAATCAT ATCAAAAAA 1380  
AAAAAAAAG G 1391

(2) INFORMATION FOR SEQ ID NO: 121:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2183 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: UCMCL5T01  
(B) CLONE: 1969948



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121 :

CTCTGTGAAC ATATGATGAG AGAAGCCAAG ATCATGCAGT ATAAGTACCT ACTGTTTCAGT 60  
 CTTACAGCCA TAGTGAAGCT TGAATCCCT CAGAACTA TTTTGGTGCA GACTTTGCTG 120  
 AGGGTGACCC AGGAACGTAT CAATGAGTGT GATGAGATAT GCCTTTCAGT TTTGTCAACT 180  
 GTTTTAGAGG CAATGGAACC ATGCAAGAAT GTTCATGTTC TACGAACGGG ATTCAGAATA 240  
 CTAGTTGATC AGCAAGTTTG GAAAATAGAA GATGTCTTCA CATTACAAGT TGTGATGAAG 300  
 TGTATTGGAA AAGATGCACC GATTGCTCTT AAGAGGAAAC TGGAGATGAA AGCCTTGAGG 360  
 GGATTAGACA GATTTTCTGT TTTGAATAGC CAACACATGT TTGAAGTACT AGCTGCCATG 420  
 AATCACCAGT CTCTTATACT CCTGGATGAA TGCAGTAAGG TGGTCCTAGA TAATATCCAT 480  
 GGGTGTCTT TAAGAATAAT GATCAACATA TTGCAGTCCT GCAAAGACCT CCAGTACCAT 540  
 AATTTGGATC TCTTCAAGGG ACTTGCAGAT TATGTGGCTG CAACTTTCGA CATCTGGAAG 600  
 TTCAGAAAAG TTCTTTTTAT CCTCATTTTA TTTGAAAACC TTGGCTTTTCG ACCTGTTGGT 660  
 TTAATGGACC TGTTTATGAA GAGAATAGTA GAGGATCCTG AATCCCTAAA CATGAAAAC 720  
 ATTCTATCTA TTCTTCATAC TTAATCTTCT CTCAATCATG TCTACAAATG CCAGAACAAA 780  
 GAACAGTTTCG TGGAAGTTAT GGCTAGTGCT CTGACTGGTT ATCTTCACAC TATTTCTTCT 840  
 GAAAACCTTAT TGGATGCAGT ATATTCATTT TGCTTGATGA ATTACTTTCC CCTGGCTCCT 900  
 TTTAATCAGC TTCTGCAAAA AGACATCATC AGTGAGCTGC TGACATCAGA TGACATGAAG 960  
 AATGCTTACA AGCTGCATAC TTTGGATACT TGTCTAAAAC TTGATGATAC TGTCTATCTG 1020  
 AGGGACATAG CCTTGTCACT CCCACAGCTG CCGCGGGAGC TGCCATCGTC ACATACAAAT 1080  
 GCAAAGGTGG CAGAGGTGCT GAGCAGCCTT CTGGGAGGTG AAGGACACTT CTCAAAGGAT 1140  
 GTGCACTTGC CACACAATTA TCATATTGAT TTTGAAATCA GAATGGACAC TAACAGGAAT 1200  
 CAAGTGCTAC CACTTTCTGA TGTGGATACA ACTTCTGCTA CAGATATTCA AAGAGTAGCT 1260  
 GTGCTATGTG TTTCCAGATC TGCTTATTGT TTGGGTTCAA GCCACCCAG AGGATTCCTT 1320  
 GCTATGAAAA TGCGGCATTT GAATGCAATG GGTTTTTCATG TGATCTTGGT CAATAACTGG 1380  
 GAGATGGACA AACTAGAGAT GGAAGATGCA GTCACATTTT TGAAGACTAA AATCTATTCA 1440  
 GTAGAAGCTC TTCCTGTTGC TGCTGTAAAT GTGCAAAGCA CACAATAAAG TGAAAATCAA 1500  
 CCTTTTCATA TTAGGAGACA TGCATTTGTA AAAATTAATA AAGATGACAA GTCAGTTGTC 1560  
 AATGGAATTG AGCTATCTGC TAAGACAAAA AATGTTACCT CAGTTCATA TTAAAAATTA 1620  
 TTTTAGGAGT GGAAGAAATG TTGTTACTGC CATTTAAAAA TATGCTGAGA AAATTCCAGA 1680  
 AGGGTTATTT TTCCAACCAC ACCTATTCCC TCTAGTGCCC AGATATTTGA TTTGTGAGCT 1740  
 GTACGTTTCA CCTTTTCATC TTTGATCTAC TAAAACTGG TTTCTTAGTT GTGAGGTGTC 1800

PF-0459 US

ACAGGCAGGT TGATGTGGGT AGTAGTCCTT GTCTTTGGAA TCTGAATATT TATACTCCTG 1860  
CTCTAAGCTG TTCTAAGACT TGGGGTTATG CCTTTAAATC ATTTTCAAGC ATTGGCCAAA 1920  
TAATAATTGG ACAAAGTTCT AAAGTTGTCA AGTGTGTAAG AATTAGTGAG GTAGCTGTTG 1980  
AAAATGAGTG AGGATGGTAT TTGTATTTGT AATAAGCACT GCAGGTAGAG ATATTTTCATG 2040  
GGTTATAATA AGAGAAACAC AGATGAGATG TAGATGGTAA GGAGTCTTAC TGTTGTTGGG 2100  
GTCCTTCCTT TCTCTTTCTT TTTTCCCCCT TACCCCTCCC ACAATTTTCAT GAAGTCTTTT 2160  
AAATTAAATA TATAGCTTNA ATT 2183

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2066 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGAST01
- (B) CLONE: 1988911

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122 :

AGAACCACTG CAGTGGAGAC TCCATGTGCA AAAGAAAAAA ACCAAATGTG AGGTCATAAA 60  
GACTTTCTGC CAGCATGTGG GTGACATTGT TTCTTTGCAG ATTTTGGCTA TGGAAAGGGG 120  
AAATGTTCTA AGCAGAGCCC CGTCAAGAGC CCACGGGACA CATTTTGGAG ATGACAGATT 180  
TGAAGATCTG GAAGAGGCAA ATCCATTCTC TTTTAGAGAG TTTCTGAAGA CCAAGAACCT 240  
CGGCCTCTCG AAAGAGGATC CGGCCAGCAG AATTTATGCA AAGGAAGCCT CGAGGCATTC 300  
CCTGGGACTT GACCACAACT CCCACCCCTC CCAAACCGGC GGGTATGGCC TGGAGTATCA 360  
GCAGCCATTT TTCGAGGATC CGACAGGGGC TGGTGACCTC CTGGATGAGG AGGAGGATGA 420  
GGACACCGGA TGGAGTGGGG CCTACCTGCC GTCCGCCATC GAGCAGACTC ACCCCGAGAG 480  
GGTCCCTGCC GGCACGTCGC CCTGCAGCAC ATACCTTTCC TTTTCTCCA CCCCGTCGGA 540  
GCTGGCAGGG CCTGAGTCTC TGCCCTCGTG GGCGTTGAGT GACACTGATT CTCGCGTGTC 600  
TCCGGCCTCT CCGGCAGGGA GTCCTAGCGC AGACTTTGCG GTTCATGGAG AGTCTCTGGG 660  
AGACAGGCAC CTGCGGACGC TGCAGATAAG TTACGACGCA CTGAAAGATG AAAATTCTAA 720  
GCTGAGAAGA AAGCTGAATG AGGTTTCAGAG CTTCTCTGAA GCTCAAACAG AAATGGTGAG 780  
GACGCTTGAG CGGAAGTTAG AAGCAAAAAT GATCAAGGAG GAAAGCGACT ACCACGACCT 840  
GGAGTCGGTG GTTCAGCAGG TGGAGCAGAA CCTGGAGCTG ATGACCAAAC GGGCTGTAAA 900

PF-0459 US

GGCAGAAAAC CACGTCGTGA AACTAAAACA GGAAATCAGT TTGCTCCAGG CGCAGGTCTC 960  
CAACTTCCAG CGAGAGAAATG AAGCCCTGCG GTGCGGCCAG GGTGCCAGCC TGACCGTGGT 1020  
GAAGCAGAAC GCCGACGTGG CCCTGCAGAA CCTCCGGGTG GTCATGAACA GTGCACAGGC 1080  
TTCCATCAAG CAACTGGTTT CCGGAGCTGA GAACTGAAT CTTGTTGCCG AAATCCTTAA 1140  
ATCTATAGAC AGAATTTCTG AAGTTAAAGA CGAGGAGGAA GACTCTTGAG GACCCCTGGG 1200  
TGTTCTCAGC ATGAAGCTCC GTGTATACCC TGAGGTCACC ACCGCTCGAT CTAAATGTGC 1260  
AGTTGTGTCC TTAAATATGC AGTCTTACC CAGAGTAAAG TGTTGATCGC AAGAGTCCAG 1320  
TGTCGTGCCC TCAGCCAGTT CTTGGCCACC ACAATGGGAG CAGCCCTGGC CGAGTTGTCT 1380  
CTGTGGTTTC TATGCAGCCC TTCTTGGCGA AATTCCTGCG ATCTTATAGA TTCTAATGAG 1440  
CTCTTGGAAG ACATTGTCAT AAAAGCCAGT GATTTTAAGA AAAAGAGTGG TTCTGGAATC 1500  
AATGTTTTCC AGTCCCATCC CAGAACATCA GTTGTAAGAT AAGTACAATT GGTGTCTCTT 1560  
GATTTTCATAA GTAGAACAAA CACTAAATGT GCCTCTGAGA TGGCCACCCC GGCAGGGGAC 1620  
CTGTGCCTTC CGCCGATGCT CAGGGCTCCC TCTGGCTCCC GGGTCACTCT TGTGGCCCCA 1680  
GTGGGTGGTC CCTGCAGTCA TGGCCTGAGT GCGCAGGGGC CACCGCGTGG CTGCTGCTGT 1740  
CCTCCTCCGG GACCCACGGG GACCAAGGTC ACACGTTCCG TGCTGTGAAG CTGTCCAGAT 1800  
GTGCCTCTTT GGCTGGGGGT TCTGGTGGAC GTTTCAAGTG GCATTTTGTA CAATGCAGGT 1860  
TAGAATTAG GAATTTCAAG TATGTGCCCC GGTCTGTCAG GTCCCAGTTG CCTTTCTGAC 1920  
GGCCCCCTC AGAGGGACGG CGATGAGCAC TAAATGCTTT TTTGACTATT TTCCTATAGA 1980  
TTTTTTTTTAA AACTTTTTTT TCCTCCTGTT CCAATTGATA GCTTTCTTAT TTAATAAATT 2040  
CTGTAGTTCA CCGCAAAAAA AAAAAA 2066

(2) INFORMATION FOR SEQ ID NO: 123:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1867 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: OVARNOT03  
(B) CLONE: 2061561

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123 :

TGGCCAGGCT GGTCTAGAAC TCCTGACTGC AAATGATCAG CCCGCCTCAG CCACCCAAAG 60  
TGTTGGGATT ACAGGTGTGA GCCACTGTGC CCAGCGTGAT TTTTTTTTTT TTAAAGCAA 120

PF-0459 US

ACTTGTCCCTT TGGTTTTGCA GAACAGGCCT GCTCCCTCTC ATCTAGCCCA TCATTTCTTG 180  
GGGCCTGAAC CCCAGTGGTC CAAAGTATTG CTTGTGAAAT TTAAAAAATG TGAATATGAT 240  
GTGGGGATGG GCCTCTTCTA CATTACCTTG GCCCAGGGGG ATCAGCTGGC TGGGAGGATT 300  
AGTGAGCACC TCTGTATTTT GAGGTCTGAG TCTTCTGGAG CTGTGTAGTT AATCTTCGGT 360  
TTCTGATAAC CCCTGGGTCC ATCTGGCCAT CAGCCTCAGC AGTGAGCAAA GCAATACCAT 420  
ACTCATTTCT ATGTTCCCTGT TCCTTCCTCT GCTCCTCCTT TGGAGAAGCA ATAATTCATG 480  
GGGGATGATA CAGTAGCACT TTACAAATGG CTCCATGTCA TTCATCCCAG GGGCCATAAT 540  
CTCTTGACC ACCTATTCTT ACTTCCTGTT CAGCTCCTTT ACAGCTTTTA TTTTCAACTG 600  
CTTCCCAACT TGGTGGGGCC TCCTTTAAGG ATGAGCCAAT AGTAAGAATG TGGCTGTAAT 660  
CAGCAGAGAC CCCTCTGAGG GGTATCTGTT CTGCAGCCCC TAGTGAAATC ATGTGATGTG 720  
AGACAGAAAC CTAAACATGG TACTTGATTC TAAACCTGTG CCAGTCTATA GCCTCTGCCT 780  
CCCCAAGCAG AGCTCAAGCC AAACGCTTCT GTCCTCTTTC CTTCTGCATT AACCCTTTGC 840  
TGATCCTCAG GGGCCACTCC CCCAACACCC CTGTACTTGG GTGAGGGATG TTGGACAGAG 900  
CCTGTTTTCA TGTACTGCAG GTGGGGGTGT GCTGACATGT TTGCTCTTGG TTGATGGAGA 960  
AGGTACAGAG GCCAGGGAGT GAAAATGGTT GACAGAAGAG GGAAGAGTTA GGTGTCTCAT 1020  
AGTCACTCAT AGTGGGGTGG TCAGGGGTAA TGGCATCTCC CCACTTTAGG CTTCTCAAAC 1080  
AGACTTTTGA CACCTCTCAA GTTCAGAGCT CTGATGTGGA AAGACAGGAG GTGTGGGGAA 1140  
GGAGGGGGAT TTCGTGTGTT TGCATGAGTG TGCGCTTCAG GCCTTGGGAG TTGGCAAGAG 1200  
GGAGGGAAGG AAGGAGAGCA AAATCTTCGG AAGGTGTTTC TTGTACCTGA GGGATCCTGC 1260  
CCTGAATCTC CATAGTCTCC ACTGTGAACT GAGGAGGGGA GGGGTGTGCT GGGGAATAAA 1320  
TCTTGTATGA GAACAATCAA AAATCAAACG AATCCCACCG ACAGACTGCT GCTCCTAGTG 1380  
ATCTGGACTC ACCTAGGGGG CATCTGGGCT GGGGTTCAN GCTTACGTNC GCGTGNATGN 1440  
GACGNCANAG CTCTTCGAAA GTGTCCCNAA ANTNCAATTC ATTGGCGGTG GTTTTAAAAG 1500  
TTCGGGCCTG GGAAACCCGG GGGNTTACCC ATTTTATCCC NCTTNGANGG CANATTCCCC 1560  
TTTTTCCCCA ATTTGGGGAA ATTTNCCAAA NGGNCCCGT AACGGTTGGC CTTTTCCCAA 1620  
AATTTNGGNC GCCCTTAATT GGGGCGATTG TGGGACCCGC GCCCTTTATA GGGGGGGGCT 1680  
TTAAAGCGGC GCNGGGGGTT CTTTGGGTGA TTACCGGCGC GGTTGACCCC GGGTAAAATA 1740  
TTGACAAGGG CCCTTTAGCG CGCGGTTTCT TGTGGGGTTT TCCTCCCAT TGGCTTTTCC 1800  
GCAAAAGTTT TGGCGGGGTT TTCCCCGAA AAGGTCTTAA AAAGCGGTGT GCCCCTCTTT 1860  
GAGGGGG 1867

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1628 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: PANCNOT04  
 (B) CLONE: 2084489

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124 :

CTCTGGGTCT GTAGCAACCG CCCAGCGTTG AGGCGCGGCT CATGCCCCCA GTATCCCGGT 60  
 CCAGCTATTC CGAGGACATC GTGGGCTCTC GGAGAAGGCG ACGCAGCTCC TCGGGGAGCC 120  
 CACCATCCCC GCAGAGCAGA TGTTCTCTTT GGGATGGCTG TTCCCGCTCT CACTCCCGCG 180  
 GCCGTGAGGG CCTCAGGCCT CCTTGGAGTG AGTTGGACGT GGGCGCTCTT TACCCCTTTA 240  
 GTCGCTCTGG GTCGCGAGGG CGGCTCCCAA GATTCCGCAA CTACGCCTTC GCGTCTCTCT 300  
 GGTGACCTC GTATAGTGGA TATCGCTACC ATCGTCACTG CTATGCAGAA GAACGGCAGT 360  
 CAGCGGAAGA CTACGAGAAG GAAGAGAGCC ATCGGCAGAG GAGGCTGAAG GAGAGAGAGA 420  
 GGATTGGGGA ATTGGGAGCG CCTGAAGTGT GGGGGCCGTC TCCAAAGTTC CCTCAGCTAG 480  
 ATTCTGACGA ACATACCCCA GTTGAGGATG AAGAAGAGGT AACGCATCAG AAAAGCAGCA 540  
 GTTCAGATTC CAACTCGGAA GAACATAGGA AAAAGAAGAC CAGTCGTTCA AGAAACAAGA 600  
 AAAAAAGAAA GAATAAGTCG TCTAAAAGAA AGCATAGGAA ATATTCTGAT AGTGACAGTA 660  
 ACTCAGAGTC TGACACAAAT TCTGACTCTG ATGATGATAA AAAGAGAGTT AAAGCCAAGA 720  
 AGAAAAAGAA GAAAAAGAAA CACAAAACAA AGAAAAAGAA GAATAAGAAA ACCAAAAAAG 780  
 AATCCAGTGA CTCAAGCTGT AAAGACTCAG AAGAGGACTT GTCAGAAGCT ACCTGGATGG 840  
 AGCAGCCAAA TGTGGCAGAT ACTATGGATT TAATAGGGCC AGAAGCACCT ATAATACATA 900  
 CCTCTCAAGA TGAAAAACCT TTGAAGTATG GCCATGCTTT GCTTCCCGGT GAAGGTGCAG 960  
 CTATGGCTGA GTATGTAAAA GCTGGAAAGC GAATCCCACG AAGAGGTGAA ATTGGGTTGA 1020  
 CAAGTGAAGA GATCGGTTCT TTTGAATGCT CAGGTTATGT CATGAGTGGT AGCAGGCATC 1080  
 GCAGAAATGGA GGCTGTACGA CTGCGTAAGG AGAACCAGAT CTACAGTGCT GATGAGAAGA 1140  
 GAGCTCTTGC ATCCTTTAAC CAAGAAGAGA GACGAAAGAG AGAAAGTAAG ATTTTAGCCA 1200  
 GTTCCGAGA GATGGTGCAC AAAAAGACAA AAGAGAAAGA TGACAAGTAA GGACTTACTT 1260  
 GTTGACAGC AGGAATTTTA ACAACAAAAA TTTTATGTGA CCAAAGTGT TAAAAGGCTT 1320  
 TACAGTGCTA CTGTAATTAC CATATTAGTA AGTCCCTCAG GAAAAAGCTT CTTTTGAGAT 1380

PF-0459 US

ATCTTTAGCA GCTTATTTTT TGTTATTTTA ACTTTAAAAA GTAATATGTG CACATGGTTT 1440  
TAAAAATATT CAACCATTAT AGGAGGAGAG TTAGTAAAAA GTGAATCTTT CACTTTAGCC 1500  
CCTGACACCT TTCCCCCAA AATATATATT TTGGTGTCTT ATATACAGAA TATACATTCT 1560  
GTGCATATAC AAGAGTATAT GTTGCAGCAT AAAGATTAAA AGCTATTAAA GTTTTTTTTC 1620  
GCTCGTTA 1628

(2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1200 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: SPLNFET02  
(B) CLONE: 2203226

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125 :

GTGGCGGCGG CGAAGGATGC ACCCGGCAGG CTTGGCGGCG GCGGCTGCGG GGACGCCCCG 60  
GCTGCCCTCG AAGCGGAGGA TCCCTGTGTC CCAGCCGGGC ATGGCCGACC CCCACCAGCT 120  
TTTCGATGAC ACAAGTTCAG CCCAGAGCCG GGGCTATGGG GCCCAGCGGG CACCTGGTGG 180  
CCTGAGTTAT CCTGCAGCCT CTCCCACGCC CCATGCAGCC TTCCTGGCTG ACCCGGTGTC 240  
CAACATGGCC ATGGCCTATG GGAGCAGCCT GGCCGCGCAG GGCAAGGAGC TGGTGGATAA 300  
GAACATCGAC CGCTTCATCC CCATCACCAA GCTCAAGTAT TACTTTGCTG TGGACACCAT 360  
GTATGTGGGC AGAAAGCTGG GCCTGCTGTT CTTCCCCTAC CTACACCAGG ACTGGGAAGT 420  
GCAGTACCAA CAGGACACCC CGGTGGCCCC CCGCTTTGAC GTCAATGCCC CGGACCTCTA 480  
CATTCCAGCA ATGGCTTTCA TCACCTACGT TTTGGTGGCT GGTCTTGCGC TGGGGACCCA 540  
GGATAGGTTT TCCCCAGACC TCCTGGGGCT GCAAGCGAGC TCAGCCCTGG CCTGGCTGAC 600  
CCTGGAGGTG CTGGCCATCC TGCTCAGCCT CTATCTGGTC ACTGTCAACA CCGACCTCAC 660  
CACCATCGAC CTGGTGGCCT TCTTGGGCTA CAAATATGTC GGGATGATTG GCGGGGTCCT 720  
CATGGGCCTG CTCTTCGGGA AGATTGGCTA CTACCTGGTG CTGGGCTGGT GCTGCGTGGC 780  
CATCTTTGTG TTCATGATCC GGACGCTGCG GCTGAAGATC TTGGCAGACG CAGCAGCTGA 840  
GGGGGTCCCG GTGCGTGGGG CCCGGAACCA GCTGCGCATG TACCTGACCA TGGCGGTGGC 900  
GGCGGCGCAG CCTATGCTCA TGTA CTGGTGGCT CACCTTCCAC CTGGTGGCGT GAGCGCGCCC 960  
GCTGAACCTC CCGCTGCTGC TGCTGCTGCT GGGGGCCACT GTGGCCGCGG AACTCATCTC 1020

PF-0459 US

CTGCCTGCAG GCCCAAGGT CCACCCTGTC TGGCCACAGG CACCGCCTCC ATCCCATGTC 1080  
CCGCCCAGCC CCGCCCCCAA CCCAAGGTGC TGAGAGATCT CCAGCTGCAC AGGCCACCGC 1140  
CCCAGGGCGT GGCCGCTGTT ACAGAAACAA TAAACCCTGA TGGGCATGGC AAAAAAAAAA 1200

(2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1093 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PROSNOT16
- (B) CLONE: 2232884

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126 :

AGAGCCCCAG CCACGCCGGC CCAGGTGGCC TCAGGTGAGG GGGGGCGGAC GCACCTGTGG 60  
GGACGGGACG ACGAGTTCAA GCCTCCGTGG GTGCAGTTGG TCGCCAGCGA GGGATGCGGA 120  
GACGCCCCTG AACGACCATG GCATCGGCCG ACGAGCTGAC CTTCCATGAA TTCGAGGAGG 180  
CCACTAATCT TCTGGCTGAC ACCCCAGATG CAGCCACCAC CAGCAGAAGC GATCAGCTGA 240  
CCCCACAAGG GCACGTGGCT GTGGCCGTGG GCTCAGGTGG CAGCTATGGA GCCGAGGATG 300  
AGGTGGAGGA GGAGAGTGAC AAGGCCGCGC TCCTGCAGGA GCAGCAGCAG CAGCAGCAGC 360  
CGGGATTCTG GACCTTCAGC TACTATCAGA GCTTCTTTGA CGTGGACACC TCACAGGTCC 420  
TGGACCGGAT CAAAGGCTCA CTGCTGCCCC GGCCTGGCCA CAACTTTGTG CGGCACCATC 480  
TGCGGAATCG GCCGGATCTG TATGGCCCCCT TCTGGATCTG TGCCACGTTG GCCTTTGTCC 540  
TGGCCGTCAC TGGCAACCTG ACGCTGGTGC TGGCCCAGAG GAGGGACCCC TCCATCCACT 600  
ACAGCCCCCA GTTCCACAAG GTGACCGTGG CAGGCATCAG CATCTACTGC TATGCGTGGC 660  
TGGTGCCCCCT GGCCCTGTGG GGCTTCCTGC GGTGGCGCAA GGGTGTCCAG GAGCGCATGG 720  
GGCCCTACAC CTTCTGGAG ACTGTGTGCA TCTACGGCTA CTCCCTCTTT GTCTTCATCC 780  
CCATGGTGGT CCTGTGGCTC ATCCCTGTGC CTTGGCTGCA GTGGCTCTTT GGGGCGCTGG 840  
CCCTGGGCCT GTCAGCCGCC GGGCTGGTAT TCACCCTCTG GCCCGTGGTC CGTGAGGACA 900  
CCAGGCTGGT GGCCACAGTG CTGCTGTCCG TGGTCGTGCT GCTCCACGCC CTCCTGGCCA 960  
TGGGCTGTAA GTTGTACTTC TTCCAGTCGC TGCCTCCGGA GAACGTGGCT CCTCCACCCC 1020  
AAATCACATC TCTGCCCTCA AACATCGCGC TGTCCCCTAC CTTGCCGCAG TCCCTGGCCC 1080  
CCTCCTAGGA AGG 1093

(2) INFORMATION FOR SEQ ID NO: 127:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1121 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: COLNNOT11  
 (B) CLONE: 2328134

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127 :

GCGGGGGATG ACGCCACGGA CATGCTGGCC GAGACCGGCG GGGTGGGGGA CGTGTGCGCG 60  
 GGCCGGGTGG CCTCGGTGCG TACCTGGGC GCGGACAGCT GCCTCATTAG TATTCGTACC 120  
 CACGAGGCGG CGCAGCGGGC CCTCGGGGAC AGCGAGCGTC GCGGCTATGG CTTATCACTC 180  
 GGGCTACGGA GCCCACGGCT CCAAGCACAG GGCCCGGGCA GCGCCGGATC CCCCTCCCCT 240  
 CTTGATGAC ACAAGCGGTG GTTATTCCAG CCAGCCCGGG GGATACCCAG CCACAGGAGC 300  
 AGACGTGGCC TTCAGTGTCA ACCACTTGCT TGGGGACCCA ATGGCCAATG TGGCTATGGC 360  
 CTATGGCAGC TCCATCGCAT CCCATGGGAA GGACATGGTG CACAAGGAGC TGCACCGTTT 420  
 TGTGTCTGTG AGCAAACCTCA AGTATTTTTT TGCTGTGGAC ACAGCCTACG TGGCCAAGAA 480  
 GCTAGGGCTG CTGGTCTTCC CCTACACACA CCAGAACTGG GAAGTGCAGT ACAGTCGTGA 540  
 TGCTCCTCTG CCCCCCGGC AAGACCTCAA CGCCCTGAC CTCTATATCC CCACGATGGC 600  
 CTTCACTACT TACGTGCTCC TGGCTGGGAT GGCACCTGGC ATTCAGAAAA GGTTCTCCCC 660  
 GGAGGTGCTG GGCCTGTGTG CAAGCACAGC GCTGGTGTGG GTGGTGATGG AGGTGCTGGC 720  
 CCTGCTCCTG GGCCTCTACC TGGCCACCGT GCGCAGTGAC CTGAGCACCT TTCACCTGCT 780  
 GGCCTACAGT GGCTACAAAT ACGTGGGAAT GATCCTCAGT GTGCTCACGG GGCTGCTGTT 840  
 CGGCAGCGAT GGCTACTACG TGGCGCTGGC CTGGACCTCA TCGGCGCTCA TGTACTTCAT 900  
 TGTGCGCTCT TTGCGGACAG CAGCCCTGGG CCGCGACAGC ATGGGGGGCC CCGTCCCCCG 960  
 GCAGCGTCTC CAGCTCTACC TGAATCTGGG AGCTGCAGCC TTCCAGCCCC TCATCATATA 1020  
 CTGGCTGACT TTCCACCTGG TCCGGTGACC CCCTGGCCCC AGATGGCACT GAGTTTTTCA 1080  
 TTCATTGAAG ATTTGATTTT CTTGAAAAAA AAAAAAAAAG G 1121

(2) INFORMATION FOR SEQ ID NO: 128:



- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1861 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: ISLTNOT01  
 (B) CLONE: 2382718

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128 :

CGCGGACTGT GTCTGTTCCC AGGAGTCCTT CGGCGGCTGT TGTGTCAGTG GCCTGATCGC 60  
 GATGGGGACA AAGGCGCAAG TCGAGAGGAA ACTGTTGTGC CTCTTCATAT TGGCGATCCT 120  
 GTTGTGCTCC CTGGCATTGG GCAGTGTTAC AGTGCACCTCT TCTGAACCTG AAGTCAGAAT 180  
 TCCTGAGAAT AATCCTGTGA AGTTGTCCTG TGCCTACTCG GGCTTTTCTT CTCCCCGTGT 240  
 GGAGTGGAAG TTTGACCAAG GAGACACCAC CAGACTCGTT TGCTATAATA ACAAGATCAC 300  
 AGCTTCCTAT GAGGACCGGG TGACCTTCTT GCCAACTGGT ATCACCTTCA AGTCCGTGAC 360  
 ACGGGAAGAC ACTGGGACAT ACACTTGTAT GGTCTCTGAG GAAGGCGGCA ACAGCTATGG 420  
 GGAGGTCAAG GTCAAGCTCA TCGTGCTTGT GCCTCCATCC AAGCCTACAG TTAACATCCC 480  
 CTCCTCTGCC ACCATTGGGA ACCGGGCAGT GCTGACATGC TCAGAACAAG ATGGTTCCCC 540  
 ACCTTCTGAA TACACCTGGT TCAAAGATGG GATAGTGATG CCTACGAATC CAAAAGCAC 600  
 CCGTGCCTTC AGCAACTCTT CCTATGTCCT GAATCCCACA ACAGGAGAGC TGGTCTTTGA 660  
 TCCCCTGTCA GCCTCTGATA CTGGAGAATA CAGCTGTGAG GCACGGAATG GGTATGGGAC 720  
 ACCCATGACT TCAAATGCTG TGCGCATGGA AGCTGTGGAG CGGAATGTGG GGGTCATCGT 780  
 GGCAGCCGTC CTTGTAACCC TGATTCTCCT GGGAATCTTG GTTTTTGGCA TCTGGTTTGC 840  
 CTATAGCCGA GGCCACTTTG ACAGAACAAA GAAAGGGACT TCGAGTAAGA AGGTGATTTA 900  
 CAGCCAGCCT AGTGCCCGAA GTGAAGGAGA ATTCAAACAG ACCTCGTCAT TCCTGGTGTG 960  
 AGCCTGGTGC GCTCACC GCC TATCATCTGC ATTTGCCTTA CTCAGGTGCT ACCGGACTCT 1020  
 GGCCCCTGAT GTCTGTAGTT TCACAGGATG CCTTATTTGT CTTCTACACC CCACAGGGCC 1080  
 CCCTACTTCT TCGGATGTGT TTTTAATAAT GTCAGCTATG TGCCCCATCC TCCTTCATGC 1140  
 CCTCCCTCCC TTTCTACCA CTGCTGAGTG GCCTGGAAC TGTTTAAAGT GTTTATTCCC 1200  
 CATTTCTTTG AGGGATCAGG AAGGAATCCT GGGTATGCCA TTGACTTCCC TTCTAAGTAG 1260  
 ACAGCAAAAA TGGCGGGGGT CGCAGGAATC TGCACTCAAC TGCCACCTG GCTGGCAGGG 1320  
 ATCTTTGAAT AGGTATCTTG AGCTTGGTTC TGGGCTCTTT CCTTGTGTAC TGACGACCAG 1380  
 GGCCAGCTGT TCTAGAGCGG GAATTAGAGG CTAGAGCGGC TGAAATGGTT GTTTGGTGAT 1440  
 GAACTGGGG TCCTTCCATC TCTGGGGCCC ACTCTCTTCT GTCTTCCCAT GGGAAGTGCC 1500

PF-0459 US

ACTGGGATCC CTCTGCCCTG TCCTCCTGAA TACAAGCTGA CTGACATTGA CTGTGTCTGT 1560  
GGAAAATGGG AGCTCTTGTT GTGGAGAGCA TAGTAAATTT TCAGAGAACT TGAAGCCAAA 1620  
AGGATTTAAA ACCGCTGCTC TAAAGAAAAG AAAACTGGAG GCTGGGCGCA GTGGCTCACG 1680  
CCTATAATCC CAGAGGCTGA GGCAGGCGGA TCACCTGAGG TCGGGAGTTC GGGATCAGCC 1740  
TGACCAACAT GGAGAAACCC TACTGAGAAT ACAAAGTTAG CCAGGCATGG TGGTGCATGC 1800  
CTGTAATCCC AGCTGCTCAG GAGCCTGGCA ACAAGAGCAA AACTCCAGCT CAAAAAAAAA 1860  
A 1861

(2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1975 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: ENDANOT01
- (B) CLONE: 2452208

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129 :

GTTTGGAGGA GACTCGGATA TACCTTCTCA GAAGCTGCAC AGGAGGAAAG CAGTGACAAA 60  
GAAAGAAGTT GTCATTCTTT GCACGAAACT GGATGGCTTC TACAGGGAGC CAGGCCTCTG 120  
ATATAGACGA GATTTTTGGA TTCTTCAACG ATGGCGAACC TCCCACCAAA AAGCCCAGGA 180  
AGCTGCTTCC AAGCTTAAAA ACTAAGAAGC CTCGAGAACT TGTGCTAGTG ATTGGAACAG 240  
GCATTAGTGC TGCAGTTGCG CCCCAGTTC CAGCCCTCAA ATCCTGGAAG GGGTTAATTC 300  
AGGCCTTACT GGATGCTGCC ATTGATTTTG ATCTTTTAGA AGATGAGGAG AGCAAAAAGT 360  
TTCAGAAATG TCTCCATGAA GACAAGAACC TGGTCCATGT TGCCCATGAC CTTATCCAGA 420  
AACTCTCTCC TCGTACCAGT AATGTTGAT CCACATTTTT CAAGGACTGT TTATATGAAG 480  
TATTTGATGA CTTGGAGTCA AAGATGGAAG ATTCTGGAAG ACAGCTACTT CAGTCAGTTC 540  
TCCACCTGAT GGAAAATGGA GCCCTCGTAT TAACTACAAA TTTTGATAAT CTCTTGGAAC 600  
TGTATGCAGC AGATCAGGGG AAACAGCTTG AATCCCTTGA CTTACTGAT GAGAAAAAGG 660  
TCCTCGAGTG GGCTCAGGAG AAGCGTAAGC TGAGCGTGTT GCATATTCAC GGAGTCTACA 720  
CCAACCCTAG TGGCATTGTC CTTATCCGG CTGGATATCA GAACGTGCTC AGGAACACTG 780  
AAGTCATGAG AGAAATTCAG AAACTCTACG AAAACAAGTC ATTTCTTTTC CTGGGCTGTG 840  
GCTGGAAGTGT GGATGACACC ACTTTCCAGG CCCTTTTCTT GGAGGCTGTC AAGCATAAAT 900

PF-0459 US

CTGACCTAGA ACATTTTCATG CTGGTTCGGA GAGGAGACGT AGATGAGTTC AAAAAGCTTC 960  
GAGAAAACAT GCTGGACAAG GGGATTAAAG TCATCTCCTA TGGAGATGAC TATGCCGATC 1020  
TTCCAGAATA TTTCAAGCGA CTGACATGTG AGATCTCCAC AAGGGGTACA TCAGCAGGGA 1080  
TGGTGAGAGA AGGTCAGCTA AATGGCTCAT CTGCAGCACA CAGTGAAATA AGAGGCTGTA 1140  
GTACATGAGC GAGCTAGAGA AATCACCACC GTTTAGACCA AGCTGTAAGG CCCTACTACA 1200  
GACAGTGTTT AACAAGTAA CTTACAAGAA CCCAACACAA TTCCCAGAAA GTAACAATAG 1260  
CCAGAGGTTG AAGGGCGGGG TAGAAGAGGG GGAATGTTG CAGCGTAATC CTTCATACCA 1320  
CCTGGTTCTT GATATTCTGC CGCCTGTTCA AGTTCAAGAA TAAAAGCGAC AGCAGGACCC 1380  
AAATGCAGCT CCCAACCAC TCCCCAGGCT AGACATGCTT GTGTCCACAC AGCACACCAA 1440  
TGTGATACTT CCACTGACCG GCTGCAGCTC TGCATGAAGG ACTCGGGGTC TGGATGCCAT 1500  
GGAATCACTG TGGCTCTTGT TGCAGTTTTG TACTCTATAC TTGGTTTTTC AATTAAGCTT 1560  
AATGGCTTTT TTAACATG ACTTGAAGCT CTAGTTTCT AGATCTTTTA CAGTGTACAG 1620  
TATTTTACAT AACTAAGCTG TATTAAGC TTGTTCACTT ACTTGCCAGG ACCCTGGCTC 1680  
TACTTTTAGA GTCATTGTAA GAACTCTAA CTTGCATCAA GGTACTAATA AGCTTAATTT 1740  
TAATAACCCA AAGTTTAAAG GTTCCGATCT TTCTCCTGG GGTGGAGTGA TCTCATTCTC 1800  
AGGACAACCG TTTACTTACC TGATTCCTCG GAGCATTATC AACTTCTGCT CTGTTGTCCT 1860  
GACCATACAT ATGTCCTAGA ACTACAGTTA AGTGTGTTGT GGAATTTTAG TTTTGAATCC 1920  
GGAATAAATG AAGTCCCAGG ACTCAAAGAA GAGAGAAAAA AAAAAAGGGG GCCCC 1975

(2) INFORMATION FOR SEQ ID NO: 130:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2160 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: ENDANOT01  
(B) CLONE: 2457825

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130 :

TCTACTGTCC CCTGCCCTGT ACCCCCAGGC ATTGATCTGG AGAACATTGT GTACTACAAG 60  
GACGACACCC ACTACTTTGT GATGACAGCC AAGAAGCAGT GCCTGCTGCG GCTGGGGGTG 120  
CTGCGCCAGG ACTGGCCAGA CACCAATCGG CTGCTGGGCA GTGCCAATGT GGTGCCCCGAG 180  
GCTCTGCAGC GCTTTACCCG GGCAGCTGCT GACTTTGCCA CCCATGGCAA GCTCGGGAAA 240

PF-0459 US

CTAGAGTTTG CCCAGGATGC CCATGGGCAG CCTGATGTCT CTGCCTTTGA CTCACGAGC 300  
ATGATGCGGG CAGAGAGTTC TGCTCGTGTG CAAGAGAAGC ATGGCGCCCC CCTGCTGCTG 360  
GGACTGGTGG GGGACTGCCT GGTGGAGCCC TTCTGGCCCC TGGGCACTGG AGTGGCACGG 420  
GGCTTCCTGG CAGCCTTTGA TGCAGCCTGG ATGGTGAAGC GGTGGGCAGA GGGCGCTGAG 480  
TCCCTAGAGG TGTGGCTGA GCGTGAGAGC CTGTACCAGC TTCTGTCACA GACATCCCCA 540  
GAAAACATGC ATCGCAATGT GGCCAGTAT GGGCTGGACC CAGCCACCCG CTACCCCAAC 600  
CTGAACCTCC GGGCAGTGAC CCCCATCAG GTACGAGACC TGTATGATGT GCTAGCCAAG 660  
GAGCCTGTGC AGAGGGACAA CGACAAGACA GATACAGGGA TGCCAGCCAC CGGGTCGGCA 720  
GGCACCAGG AGGAGCTGCT ACGCTGGTGC CAGGAGCAGA CAGCTGGGTA CCCGGGAGTC 780  
CACGTCTCCG ATTTGTCTTC CTCCTGGGCT GATGGGCTAG CTCTGTGTGC CCTGGTGTAC 840  
CGGCTGCAGC CTGGCCTGCT GGAACCCCTCA GAGCTGCAGG GGCTGGGAGC TCTGGAAGCA 900  
ACTGCTTGGG CACTAAAGGT GGCAGAGAAT GAGCTGGGCA TCACACCGGT GGTGTCTGCA 960  
CAGGCCGTGG TAGCAGGGAG TGACCCACTG GGCCTCATTG CCTACCTCAG CCACTTCCAC 1020  
AGTGCCTTCA AGAGCATGGC CCACAGCCCA GGCCCTGTCA GCCAGGCCTC CCCAGGGACC 1080  
TCCAGTGCTG TATTATTCCT TAGTAAACTT CAGAGGACCC TGCAGCGATC CCGGGCCAAG 1140  
GAAAATGCAG AGGATGCTGG TGGCAAGAAG CTGCGCTTGG AGATGGAGGC CGAGACCCCA 1200  
AGTACTGAGG TGCCACCTGA CCCAGAGCCT GGTGTACCCC TGACACCCCC ATCCCAACAC 1260  
CAGGAGGCCG GTGCTGGGGA CCTGTGTGCA CTTTGTGGGG AACACCTCTA TGTCTGGAA 1320  
CGCCTCTGTG TCAACGGCCA TTTCTTCCAC CGGAGCTGCT TCCGCTGCCA TACCTGTGAG 1380  
GCCACACTGT GGCCAGGTGG CTACGAGCAG CACCAGGCA GTAGAACGTC TCAGTTCTTC 1440  
TTCTCAGCTC TTGTGGCCAT GGAGAAGGAG GAAAAAGAGA GTCCCTTCTC CAGTGAAGAG 1500  
GAAGAAGAAG ATGTGCCTTT GGACTCAGAT GTGGAACAGG CCCTGCAGAC CTTTGCCAAG 1560  
ACCTCAGGCA CCATGAATAA CTACCCAACA TGCGCTCGGA CTCTGCTGCG CCGTGCGAAG 1620  
GAGGAGGAGA TGAAGAGGTT CTGCAAGGCC CAGACCATCC AACGGCGACT AAATGAGATT 1680  
GAGGCTGCCT TGAGGGAGCT AGAGGCCGAG GGCCTGAAGC TGGAGCTGGC CTTGAGGCGC 1740  
CAGAGCAGTT CCCCAGAACA GCAAAAGAAA CTATGGGTAG GACAGCTGCT ACAGCTCGTT 1800  
GACAAGAAAA ACAGCCTGGT GGCTGAGGAG GCCGAGCTCA TGATCACGGT GCAGGAATTG 1860  
AATCTGGAGG AGAAACAGTG GCAGCTGGAC CAGGAGCTAC GAGGCTACAT GAACCGGGAA 1920  
GAAAACCTAA AGACAGCTGC TGATCGGCAG GCTGAGGACC AGGTCCTGAG GAAGCTGGTG 1980  
GATTTGGTCA ACCAGAGAGA TGCCCTCATC CGCTTCCAGG AGGAGCGCAG GCTCAGCGAG 2040  
CTGGCCTTGG GGACAGGGGC CCAGGGCTAG ACGAGGGTGG GCCGTCTGCT TTCGTTCCCA 2100

CAAAGAAAGC ACCTCACCCC AGCACAGTGC CACCCCTGTT CATCTGGGCT GCCTGGCAGA 2160

(2) INFORMATION FOR SEQ ID NO: 131:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 546 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: THP1NOT03  
(B) CLONE: 2470740

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131 :

GAGGAAGAAG AGGAAGAGGG GGCTCCGATT GGGACCCCTA GGGATCCTGG AGATGGTTGT 60  
CCTTCCCCCG ACATCCCTCC TGAACCCCT CCAACACACC TGAGGCCCTG CCCTGCCAGC 120  
CAGCTCCCTG GACTCCTGTC CCATGGCCTC CTGGCCGGCC TCTCCTTTGC AGTGGGGTCC 180  
TCCTCTGGCC TCCTGCCCCT CCTGCTGCTG CTGCTGCTTC CATTGCTGGC AGCCCAGGGT 240  
GGGGGTGGCC TGCAGGCAGC GCTGCTGGCC CTTGAGGTGG GGCTGGTGGG TCTGGGGGCC 300  
TCCTACCTGC TCCTTTGTAC AGCCCTGCAC CTGCCCTCCA GTCTTTTCCT ACTCCTGGCC 360  
CAGGGTACCG CACTGGGGGC CGTCCTGGGN CATGAGCTGG CGCCGAAGGC TCATGGGTGT 420  
TCCCCTGGGG CTTTGGAAC TGCCTGGTTCT TAAGCTTNGG CAAGGCCTAG CTCCAACCTC 480  
TGGTGGCTAA TGGCANCCGG GGGGGAANAT GGGTTCNGGA AAAAGGGCCC CCGGGTTTCA 540  
CCGGGG 546

(2) INFORMATION FOR SEQ ID NO: 132:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 581 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: SMCANOT01  
(B) CLONE: 2479092

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132 :

GCCATGGAGG CCCTGAGGAG GGCCACGAG GTCGCGCTCC GCCTGCTGCT GTGTAGGCCG 60  
TGGGCCTCGC GCGCCGCCGC CCGCCCAAG CCCAGCGCCT CGGAGGTGCT GACGCGGCAT 120

PF-0459 US

CTGCTGCAGC GGCGCCTGCC GCACTGGACC TCCTTCTGCG TGCCCTACAG CGCCGTCCGC 180  
AACGACCAGT TCGGCCTCTC GCACTTCAAC TGGCCGGTGC AGGGCGCCAA CTACCACGTC 240  
CTGCGCACCG GCTGCTTCCC CTTTCATCAAG TACCACTGCT CCAAGGCTCC CTGGCAGGAC 300  
CTGGCCCCGGC AGAACCGCTT CTTACGGCG CTCAAGGTCG TCAACCTCGG TATTCCAAC 360  
TTATTATATG GACTTGGCTC CTGGTTATTT GCCAGAGTCA CAGAGACTGT GCATACCAGT 420  
TATGGACCCA TAACAGTTTA TTTTCTCAAT AAAGAAGATG AAGGTGCCAT GTATTGAAAG 480  
TGTGCGTCAA AGAACATAAA TATCAGTGGA TTTTCTCTGT GTATATGTGC AGTATTTATT 540  
TTGATCCTT TAAATAAAA CTTTGCAAA TAAAAAAAAA A 581

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1259 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: SMCANOT01  
(B) CLONE: 2480544

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133 :

GGGCTGGGCC CCGCCGAGC TCCAGCTGGC CGGCTTGGTC CTGCGGTCCC TTCTCTGGGA 60  
GGCCCGACCC CGGCCGCGCC CAGCCCCAC CATGCCACCC GCGGGGCTCC GCCGGGCGC 120  
GCCGCTCACC GCAATCGCTC TGTTGGTGCT GGGGGCTCCC CTGGTGCTGG CCGGCGAGGA 180  
CTGCCTGTGG TACCTGGACC GGAATGGCTC CTGGCATCCG GGGTTTAACT GCGAGTTCTT 240  
CACCTTCTGC TGCGGGACCT GCTACCATCG GTACTGCTGC AGGGACCTGA CCTTGCTTAT 300  
CACCGAGAGG CAGCAGAAGC ACTGCCTGGC CTTAGCCCC AAGACCATAG CAGGCATCGC 360  
CTCAGCTGTG ATCCTCTTTG TTGCTGTGGT TGCCACCACC ATCTGCTGCT TCCTCTGTTT 420  
CTGTTGCTAC CTGTACCGCC GGCGCCAGCA GCTCCAGAGC CCATTTGAAG GCCAGGAGAT 480  
TCCAATGACA GGCATCCCAG TGCAGCCAGT ATACCCATAC CCCCAGGACC CCAAAGCTGG 540  
CCCTGCACCC CCACAGCCTG GCTTCATGTA CCCACCTAGT GGTCCTGCTC CCCAATATCC 600  
ACTCTACCCA GCTGGGCCCC CAGTCTACAA CCCTGCAGCT CCTCCTCCCT ATATGCCACC 660  
ACAGCCCTCT TACCCGGGAG CCTGAGGAAC CAGCCATGTC TCTGCTGCCC CTTAGTGAT 720  
GCCAACCTTG GGAGATGCCC TCATCCTGTA CCTGCATCTG GTCCTGGGGG TGGCAGGAGT 780  
CCTCCAGCCA CCAGGCCCCA GACCAAGCCA AGCCCTGGGC CCTACTGGGG ACAGAGCCCC 840

PF-0459 US

AGGGAAGTGG AACAGGAGCT GAACTAGAAC TATGAGGGGT TGGGGGGAGG GCTTGGGAATT 900  
ATGGGCTATT TTTACTGGGG GCAAGGGAGG GAGATGACAG CCTGGGTCAC AGTGCCTGTT 960  
TTCAAATAGT CCCTCTGCTC CCAAGATCCC AGCCAGGAAG GCTGGGGCCC TACTGTTTGT 1020  
CCCCTCTGGG CTGGGGTGGG GGGAGGGAGG AGGTTCCGTC AGCAGCTGGC AGTAGCCCTC 1080  
CTCTCTGGCT GCCCCACTGG CCACATCTCT GGCTGCTAG ATTAAAGCTG TAAAGACATA 1140  
ACTCATATCA GTCGCATCAT TGGACCCATC CACACCTTCC AGGAACACCG NCTTCAGCTG 1200  
GGCCCAGACT GTTGCCCACT CCATATTCCA AAAGTAGGGG AGGGCCAGCA CCAGCATCG 1259

(2) INFORMATION FOR SEQ ID NO: 134:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2033 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAITUT21
- (B) CLONE: 2518547

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134 :

CGGCTCGAGG CCGCAGCCCC ATGGACAGTC TTCTGCACCC CCGGGAGCGC CCTGGATCCA 60  
CTGCCTCCGA GAGCTCAGCC TCTCTGGGCA GTGAGTGGGA CCTCTCAGAA TCTTCTCTCA 120  
GCAACCTGAG TCTTCGCCGT TCCTCAGAGC GCCTCAGTGA CACCCTGGA TCCTTCCAGT 180  
CACCTTCCCT GGAAATTCTG CTGTCCAGCT GCTCCCTGTG CCGTGCCTGT GATTCGCTGG 240  
TGTATGATGA GGAAATCATG GCTGGCTGGG CACCTGATGA CTCTAACCTC AACACAACCT 300  
GCCCCTTCTG CGCCTGCCCC TTTGTGCCCC TGCTCAGTGT CCAGACCCTT GATTCCCGGC 360  
CCAGTGTCCT CAGCCCCAAA TCTGCTGGTG CCAGTGGCAG CAAAGATGCT CCTGTCCCTG 420  
GTGGTCCTGG CCCTGTGCTC AGTGACCGAA GGCTCTGCCT TGCTCTGGAT GAGCCCAGCT 480  
CTGCAACGGG CACATGGGGG GAGCCTCCCG GCGGGTTGAG AGTGGGGCAT GGGCATACT 540  
GAGCCCCCTG GTGCTGCGTA AGGAGCTGGA GTCGCTGGTA GAGAACGAGG GCAGTGAGGT 600  
GCTGGCGTTG CCTGAACTGC CCTCTGCCCA CCCCATCATC TTCTGGAACC TTTTGTGGTA 660  
TTTCCAACGG CTACGCCTGC CCAGTATTCT ACCAGGCTG GTGCTGGCCT CCTGTGATGG 720  
GCCTTCGCAC TCCCAGGCCC CATCTCCTTG GCTAACCCCT GATCCAGCCT CTGTTTCAGGT 780  
ACGGCTGCTG TGGGATGTAC TGACCCTGA CCCCAATAGC TGCCCACCTC TCTATGTGCT 840  
CTGGAGGGTC CACAGCCAGA TCCCCAGCG GGTGGTATGG CCAGGCCCTG TACCTGCATC 900

CCTTAGTTTG GCACTGTTGG AGTCAGTGCT GCGCCATGTT GGACTCAATG AAGTGCACAA 960  
 GGCTGTGGGG CTCCTGCTGG AAACCTCTAGG GCCCCACCC ACTGGCCTGC ACCTGCAGAG 1020  
 GCGAATCTAC CGTGAGATAT TATTCCTGAC AATGGCTGCT CTGGGCAAGG ACCACGTGGA 1080  
 CATAGTGGCC TTCGATAAGA AGTACAAGTC TGCCTTTAAC AAGCTGGCCA GCAGCATGGG 1140  
 CAAGGAGGAG CTGAGGCACC GCGGGGCGCA GATGCCCCACT CCCAAGGCCA TTGACTGCCG 1200  
 AAAATGTTTT GGAGCACCTC CAGAATGCTA GAGACCTTAA GCTTCCCTCT CCAGCCTAGG 1260  
 GTGGGGAAGT GAGGAAGAAG GGATTCTAGA GTTAAACTGC CTCCCTGTTG CCTTCATGGA 1320  
 GTTGGGAACA GGCTGGGAAG GATGCCCCAGT CAAAGGCTCC AAGCGAGGAC AACAGGAAGA 1380  
 GGGATCCACT GTTACCAAAA GTCCTGATTC CCCCATCACC AACCTACCCA GTTTGTTCGT 1440  
 GCTGATGTTG GGGGAGATCT GGGGGGAGTT GGTACAGCTC TGTTCTTCCC TTGTCCTATA 1500  
 CCGGGAAGTC CCCTCCAGGG TACCCACAGA TCTGCATTGC CCTGGTCATT TTAGAAGTTT 1560  
 TTGTTTTTAAA AAACAACCTGG AAAGATGCAG AGCTACTGAG CCTTTGCCCT GAATGGGAGG 1620  
 TAGGGATGTC ATTCTCCACC AATAATGGTC CCTCTTCCCT GACGTTGCTG AAGGAGCCCA 1680  
 AGGCTCTCCA TGCCTTTCTA CCTAAGTGTT TGTATTTTAT TTAAATTAT TTATTCTGGA 1740  
 GCCACAGCCC CCTTGCTTAT GAGGTTCTTA TGGAGAGTGA GAAAGGGAAG GGAAATAGGG 1800  
 CACCATGGTC CGGTGGTTTG TAGTTCCTTC AAAGTCAGGC ACTGGGAGCT AGAGGAGTCT 1860  
 CAAGCTCCCC TTAGGAAGAA CTGGTGCCCC CTCCAGTCCT AATTTTCTT GCCTGCCCCG 1920  
 CCTTGGGGAA TGCCTCACCC ACCCAGGTCC TGACCTGTGC AATAAGGATT GTCCCTGCG 1980  
 AAGTTTTGTT GGATGTAAAT ATAGTAAAAG CTGCTTCTGT CTTTTTCAAA AAA 2033

(2) INFORMATION FOR SEQ ID NO: 135:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3007 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:
  - (A) LIBRARY: GBLANOT02
  - (B) CLONE: 2530650

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135 :

GCCCACTGGG CTCTCCCGGC TGCAGTGCCA GGGCGCAGGA CGCGGCCGAT CTCCCGCTCC 60  
 CGCCACCTCC GCCACCATGC TGCTCCCCCA GCTCTGCTGG CTGCCGCTGC TCGCTGGGCT 120  
 GCTCCCGCCG GTGCCCGCTC AGAAGTTCTC GGCGCTCACG TTTTGTGAGAG TGGATCAAGA 180



PF-0459 US

TAAAGACAAG GATTGTAGCT TGGACTGTGC GGGTTCGCCC CAGAAACCTC TCTGCGCATC 240  
TGACGGAAGG ACCTTCCTTT CCCGTTGTGA ATTTCAACGT GCCAAGTGCA AAGATCCCCA 300  
GCTAGAGATT GCATATCGAG GAAACTGCAA AGACGTGTCC AGGTGTGTGG CCGAAAGGAA 360  
GTATACCCAG GAGCAAGCCC GGAAGGAGTT TCAGCAAGTG TTCATTCTTG AGTGCAATGA 420  
CGACGGCACC TACAGTCAGG TCCAGTGTCA CAGCTACACG GGATACTGCT GGTGCGTCAC 480  
GCCCAACGGG AGGCCCATCA GCGGCACTGC CGTGGCCAC AAGACGCCCC GGTGCCCGGG 540  
TTCCGTAAAT GAAAAGTTAC CCCAACGCGA AGGCACAGGA AAAACAGATG ATGCCGCAGC 600  
TCCAGCGTTG GAGACTCAGC CTCAAGGAGA TGAAGAAGAT ATTGCATCAC GTTACCCTAC 660  
CCTTTGGACT GAACAGGTTA AAAGTCGGCA GAACAAAACC AATAAGAATT CAGTGTATC 720  
CTGTGACCAA GAGCACCAGT CTGCCCTGGA GGAAGCCAAG CAGCCCAAGA ACGACAATGT 780  
GGTGATCCCT GAGTGTGCGC ACGGCGGCCT CTACAAGCCA GTGCAGTGCC ACCCCTCCAC 840  
GGGGTACTGC TGGTGCCTCC TGGTGGACAC GGGGCGCCCC ATTCCCGGCA CATCCACAAG 900  
GTACGAGCAG CCGAAATGTG ACAACACGGG CCAGGGCCCA CCCAGCCAAA GCCCGGGACC 960  
TGTACAAGGG CCGCCAGCTA CAAGGTTGTC CGGGTGCCAA AAAGCATGAG TTTCTGACCA 1020  
GCGTTCTGGA CGCGCTGTCC ACGGACATGG TCCACGCCGC CTCCGACCCC TCCTCCTCGT 1080  
CAGGCAGGCT CTCAGAACCC GACCCAGCC ATACCCTAGA GGAGCGGGTG GTGCACTGGT 1140  
ACTTCAAAC ACTGGATAAA AACTCCAGTG GAGACATCGG CAAAAGGAA ATCAAACCCT 1200  
TCAAGAGGTT CCTTCGCAA AAATCAAAGC CCAAAAAATG TGTGAAGAAG TTTGTTGAAT 1260  
ACTGTGACGT GAATAATGAC AAATCCATCT CCGTACAAGA ACTGATGGGC TGCTGGGCG 1320  
TGGCGAAAGA GGACGGCAA GCGGACACCA AGAAACGCCA CACCCCAAGA GGTGATGCTG 1380  
AAAGTACGTC TAATAGACAG CCAAGGAAAC AAGGATAAAT GGCTCATACC CCGAAGGCAG 1440  
TTCCTAGACA CATGGGAAAT TTCCCTCACC AAAGAGCAAT TAAGAAAACA AAAACAGAAA 1500  
CACATAGTAT TTGCACTTTG TACTTTAAAT GTAAATTCAC TTTGTAGAAA TGAGCTATTT 1560  
AAACAGACTG TTTAATCTG TGAAAATGGA GAGCTGGCTT CAGAAAATTA ATCACATACA 1620  
ATGTATGTGT CCTCTTTTGA CCTTGGAAAT CTGTATGTGG TGGAGAAGTA TTTGAATGCA 1680  
TTTAGGCTTA ATTTCTTCGC CTTCACATG TTAACAGTAG AGCTCTATGC ACTCCGGCTG 1740  
CAATCGTATG GCTTTCTCTA ACCCCTGCAG TCACTTCCAG ATGCCTGTGC TTACAGCATT 1800  
GTGGAATCAT GTTGAAGCT CCACATGTCC ATGGAAGTTT GTGATGTACG GCCGACCCTA 1860  
CAGGCAGTTA ACATGCATGG GCTGGTTTGT TTCTTGGGAT TTTCTGTTAG TTTGTCTTGT 1920  
TTTGCTTTCC AGAGATCTTG CTCATACAAT GAATCACGCA ACCACTAAAG CTATCCAGTT 1980  
AAGTGCAGGT AGTCCCCTG GAGGAAATAA TATTTTCAA CTGTCGTTGG TGTGATACTT 2040

PF-0459 US

TGGCTCAAAG GATCTTTGCT TTTCCATTTT AAGCTTCTGT TTTGAGTTTT GCCCTGGGGC 2100  
TTGAATGAGT CCCAGAGAGT CGTTCGGATG GTGGGAGGCT GCCTAGGAGG CAGTAAATCC 2160  
AGTCACAGTG CCTGGGAGGG GCCCATCCTT CCAAAATGTA AATCCAGTCG CGGTGTGACC 2220  
GAGCTGGCTA ACAGGCTTGT CTGCCTGGTT TTCCTCCTAC ACGTGGACAT TATTCTCCTG 2280  
ATCCTCCTAC CTGGTCCACC CCAGGGCTAC CGGAAGGTAA AATCTTCACC TGAACCAATT 2340  
ATGAGCAGTC TCCTTACTGA AGGTACAGCC GGATACGTGG TGCCCCCGGG GCTGGTGTG 2400  
GCAGCCGGGG GGAGGTGCCT GAGGGTCCCC ACGGTTCCTT TCTGCTTTTC TGAATGCATC 2460  
AAGGGTACGA GAACTTGCCA ATGGGAAATT CATCCGAGTG GCACTGGCAG AGAAGGATAG 2520  
GAGTGGAATG CCCACACAGT GACCAACAGA ACTGGTCTGC GTGCATAACC AGCTGCCACC 2580  
CTCAGGCCTG GGCCCCAGAG CTCAGGGCAC CCAGTGTCTT AAGGAACCAT TTGGAGGACA 2640  
GTCTGAGAGC AGGAACTTCA AGCTGTGATT CTATCTCGGC TCAGACTTTT GGTGGGAAA 2700  
AGATCTTCAT GGCCCCAAAT CCCCTGAGAC ATGCCTTGTA GAATGATTTT GTGATGTTGT 2760  
GATGCTTGTG GAGCATCGCG TAAGGCTTCT TGCTTATTTA AACTGTGCAA GGTAAAAATC 2820  
AAGCCTTTGG AGCCACAGAA CCAGCTCAAG TACATGCCAA TGTTGTTTAA GAAACAGTTA 2880  
TGATCCTAAA CTTTTTGAT AATCTTTTAT ATTTCTGACC TTTGAATTTA ATCATTGTTC 2940  
TTAGATTAAA ATAAAATATG CTATTGAAAC TAAAAAATAA AAAGAGGGGA GAAGAAAAAA 3000  
AAAAAGG 3007

(2) INFORMATION FOR SEQ ID NO: 136:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1229 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: THYMNOT04
- (B) CLONE: 2652271

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136 :

CTCTCTGCTC CGGTGCAGGC CCGCAGGCGC CCTGGGCTGG GAGCAACGCG ACTGACCGTG 60  
GTCGTGGGCG GACGGCGGCT GCAGCGTGGA GGAGCTGGGG TCGCTGTGGG TCGCGAACAG 120  
AGCCCGGGAC GTGCGCGCTT GGTGCACGAT CCTGAAGGGG AGCTCCGAGG GGCCCGGGTC 180  
TCCAGGGCTG CTGCGGCCAT TCCCGGAGCC CGGCGCGGGG CCCGCGAGAT ACTGGTTTAG 240  
GCCGTCCCAG GGCTCCGGGC GCACCCGGTG GCCGCTGCTG CAGCGGAGGG AGCGCGGCGG 300

PF-0459 US

CGCGGGGGCT CGGAGACAGC GTTCTCCCG GAAGTCTTCC TCGGGCAGCA GGTGGGAAGT 360  
GGGAGCCGGA GCGGCAGCTG GCAGCGTTCT CTCCGCAGGT CGGCACCATG CGCCCTGCAG 420  
CCCTGCGCGG GGCCCTGCTG GGCTGCCTCT GCCTGGCGTT GCTTTGCCTG GCGGGTGCAG 480  
ACAAGCGCCT GCGTGACAAC CATGAGTGGA AAAAATAAT TATGGTTCAG CACTGGCCTG 540  
AGACAGTATG CGAGAAAATT CAAAACGACT GTAGAGACCC TCCGGATTAC TGGACAATAC 600  
ATGGACTATG GCCCGATAAA AGTGAAGGAT GTAATAGATC GTGGCCCTTC AATTTAGAAG 660  
AGATTAAGGA TCTTTTGCCA GAAATGAGGG CATACTGGCC TGACGTAATT CACTCGTTTC 720  
CCAATCGCAG CCGCTTCTGG AAGCATGAGT GGGAAAAGCA TGGGACCTGC GCCGCCAGG 780  
TGGATGCGCT CAACTCCAG AAGAAGTACT TTGGCAGAAG CCTGGAATC TACAGGGAGC 840  
TGGACCTCAA CAGTGTGCTT CTAAATTGG GGATAAAACC ATCCATCAAT TACTACCAAG 900  
TTGCAGATTT TAAAGATGCC CTGTCAGAG TATATGGAGT GATACCCAAA ATCCAGTGCC 960  
TTCCACCAAG CCAGGATGAG GAAGTACAGA CAATTGGTCA GATAGAACTG TGCCTACTA 1020  
AGCAAGACCA GCAGCTGCAA AACTGCACCG AGCCGGGGGA GCAGCCGTCC CCCAAGCAGG 1080  
AAGTCTGGCT GGCAAATGGG GCCGCCGAGA GCCGGGTCT GAGAGTCTGT GAAGATGGCC 1140  
CAGTCTTCTA TCCCCACCT AAAAAGACCA AGCATTGATG CCCAAGTTTT GGAAATATTC 1200  
TGTTTTAAAA AGCATGAGGT AGGCATGTC 1229

(2) INFORMATION FOR SEQ ID NO: 137:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1972 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: LUNGTUT11  
(B) CLONE: 2746976

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137 :

ACAGGGGCTT CCCCTTCGCC GCCGCCGCCG CCGCCGGCCA AGCTCCGCCG CGCCCGCGGC 60  
CCGCGGCCGC CATGCAGTTT ATGTTGCTTT TTAGTCGTCA GGGAAAGCTT CGACTGCAAA 120  
AATGGTATGT CCCACTATCA GACAAAGAGA AGAGAAAGAT CACAAGAGAA CTTGTTTCTA 180  
CCGTTTTAGC ACGGAAACCT AAAATGTGCA GCTTCCTTGA GTGGCGAGAT CTGAAGATTG 240  
TTTACAAAAG ATATGCTAGT CTGTATTTTT GCTGTGCTAT TGAGGATCAG GACAATGAAC 300  
TAATTACCCT GGAAATAATT CATCGTTATG TGGAATTACT TGACAAGTAT TTCGGCAGTG 360

PF-0459 US

TCTGTGAACT AGATATCATC TTTAATTTTG AGAAGGCTTA TTTTATTTTG GATGAGTTTC 420  
TTTTGGGAGG GGAAGTTCAG GAAACATCCA AGAAAAATGT CCTTAAAGCA ATTGAGCAGG 480  
CTGATCTACT GCAGGAGGAT GCGAAAGAAG CTGAAACCCC ACGTAGTGTT CTTGAAGAAA 540  
TTGGACTGAC ATAACCTCTCC TCCCTTGTTG ATGACTTCTT GTGGCATTTC ACACACTGTA 600  
GATGGTCACT CCCTTCATGT CCATGTTAGC TCATGGTGTA AGATGATGTC TTGTCAGTAT 660  
TACTGTTTTG CTAAGCCGCT TCATTCATGC CTACACAATT TTTTTTTAAA AGGGAACCTT 720  
AGTTAATTAA GTGATAAGGG ACTTAAATAT GAATTAGAAT GGTGCAGAAA GAGATACCTT 780  
TTCTGGATAT TTTAAAGTTT AAAGGTCAGT TTCTCTTAAT CTGATTATGT GCACATATGA 840  
AAATGGCACA TCATATACAT GTAAAATCAG GCAGTATACA TTTATTAATT ACTGTATTTG 900  
ACAAAGGAAA CTCTTAAATT ATAATGTGAA ACCTGGTTTT ATGAAACCAA AGACTAGTGC 960  
AGCATTTTCAG CATATGTAAA AAAAAAAAAA AAGGGAATTG ACATGTCACA TATCAAATGA 1020  
ATGGAACTT TGTGAACT TAAAAAGCA AATTTACTCC AAAGACTTGT ATTGGAAATT 1080  
ACATACCTTT TTTTTTTTTT TTTAAAGGAC TACAGATTAT TTTAATGAC TAAATTGGAG 1140  
TGATACTTCT TACACTAAAA ATTATTTCTT AGGCATTCTG AATCTGGGAT GAGAAACAGG 1200  
ATTGTTTCAC AATAGTAAGC ACATAATTTT TAAGGCCAAG GCACATTTGA CTCCTGAGAT 1260  
GAATTTTTTG TGGTCATAAT CAAATACTTA GTTGTTTTTG ATGCCCCAAA ATAAAGTGAG 1320  
AATGGTAATT TGCCAGGAAT TCTTCATAAC AGTATCTTAC AAAAAACGTG TTGCTCTCTT 1380  
CACAGTATTA TGTGTAAAGT CATTGTTTAA AGCACGAATG TTCCCTCTGG GGTACTTGTT 1440  
AAAGCTAAAT TTATTTTGCT TCCCTCCACT TAGAAGTGCT GCACACTTTA CAGCAGCTTC 1500  
CTTCTTTTCC ATGGCACTGC CTAGTTAACA GAAGTCTTAT AAAAATTTAA AAAGACACAT 1560  
TTCTTACAAA AAAGAGTTGA ATGAGGTAAA ATGGCATTAG ATGGCTCTAT ATTTTTTAAA 1620  
GCTATGTAAT TGTTTCAGCGT CACTTTTCTA AGTACTTATA CATATCTAAA CATGTCTTCA 1680  
TGGTTTATAT TTTCACCTAT ATATGCTGGG CTGGATTAAG CTTTGTTGTG ATTGTGACCA 1740  
ACATTCAGGC CACGTGAGCA CTGTCTTATC ACATCGCCAA TTAGTTGTAA TAAACGTTCA 1800  
ACGTACAAAC ACTGGAGTGT GTTTTTATCT CTTTCCAAAA GTTGTCAAAA CTATGCAGAG 1860  
CTGCTGAAGG AAGAATTTCT CATTTTTTTT TCAGTAAAAT GTTGAAAATT CCCCTCCATT 1920  
TGAATATGGT GGTGTGTTATA AGCACACACA AGATACATGG TGGAAGATCT AG 1972

(2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:

PF-0459 US

(A) LENGTH: 1741 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: THP1AZS08  
(B) CLONE: 2753496

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138 :

CGGGTTCCGG GCTCCGGGCT CTGGGTGGCG GCGGCTGTGA GCNGCGGCTG ANCCNCCGCG 60  
CTGCGCANC GACGCGGAAT GAAGCGGGCG CTGGGCAGGC GAAAGGGCGT GTGGTTGCGC 120  
CTGAGGAAGA TACTTTTCTG TGTTTTGGGG TTGTACATTG CCATTCCATT TCTCATCAAA 180  
CTATGTCCTG GAATACAGGC CAAACTGATT TTCTTGAATT TCGTAAGAGT TCCCTATTTT 240  
ATTGATTTGA AAAAACCACA GGATCAAGGT TTGAATCACA CGTGTAAC TAACCTGCAG 300  
CCAGAGGAAG ACGTGACCAT TGGAGTCTGG CACACCGTCC CTGCAGTCTG GTGGAAGAAC 360  
GCCCCAAGCA AAGACCAGAT GTGGTATGAG GATGCCTTGG CTTCCAGCCA CCCTATCATT 420  
CTGTACCTGC ATGGGAACGC AGGTACCAGA GGAGGCGACC ACCGCGTGGA GCTTTACAAG 480  
GTGCTGAGTT CCCTTGTTA CCATGTGGTC ACCTTTGACT ACAGAGGTTG GGGTGACTCA 540  
GTGGGAACGC CATCTGAGCG GGGCATGACC TATGACGCAC TCCACGTTTT TGACTGGATC 600  
AAAGCAAGAA GTGGTGACAA CCCCCTGTAC ATCTGGGGCC ACTCTCTGGG CACTGGCGTG 660  
GCGACAAATC TGGTGCGGCG CCTCTGTGAG CGAGAGACGC CTCCAGATGC CCTTATATTG 720  
GAATCTCCAT TACTAATAT CCGTGAAGAA GCTAAGAGCC ATCCATTTTC AGTGATATAT 780  
CGATACTTCC CTGGGTTTGA CTGGTTCTTC CTTGATCCTA TTACAAGTAG TGAATTTAAA 840  
TTTGCAAATG ATGAAAACGT GAAGCACATC TCCTGTCCCC TGCTCATCCT GCACGCTGAG 900  
GACGACCCGG TGGTGCCCTT CCAGCTTGGC AGAAAGCTCT ATAGCATCGC CGCACCAGCT 960  
CGAAGCTTCC GAGATTTCAA AGTTCAGTTT GTGCCCTTTC ATTACAGACT TGGCTACAGG 1020  
CACAAATACA TTTACAAGAG CCCTGAGCTG CCACGGATAC TGAGGGAATT CCTGGGGAAG 1080  
TCGAGCCTG AGCACCAGCA CTGAGCCTGG CCGTGGGAAG GAAGCATGAA GACCTCTGCC 1140  
CTCCTCCCGT TTTCTTCCAG TCAGCAGCCC GGTATCCTGA AGCCCCGGGG GGCCGGCACC 1200  
TGCAATGCTC AGGAGCCCAG CTCGCACCTG GAGAGCACCT CAGATCCCAG GTGGGGAGGC 1260  
CCCTGCAGGC CTGCAGTGCC CGGAGGCCCTG AGCATGGCTG TGTGGAAAGC GTGGGTGGCA 1320  
GGCATGTGGC TCTCCTTGCC GCCCCTCAAC CTGAGATCTT GTTGGGAGAC TTAATGGCAG 1380  
CAGGCAGCCA TACTGCCTG GTTGATGCTG CACTGAGCTG GACAGGGGGA GTCCGGGCAG 1440  
GGGACTCTTG GGGCTCGGGA CCATGCTGAG CTTTTTGGCA CCACCCACAG AGAACGTGGG 1500

PF-0459 US

GTCCAGGTTT TTTCTGCACC TTCCCAGCAC ATGCAGAATG ACTCCAGTGG TTCCATCGTC 1560  
CCCTCCTGCC CTGTGTACCT GCTTGCTTTT CTCAGCTGCC CCACCTCCCC TGGGCTGGCC 1620  
CACTCACCCA CAGTGAAGT GCCCGGGATC TGCACTTCCT CCCCTTTCAC CTACCTGTAC 1680  
ACCTAACCTG GCCTTAGACT GAGCTTTATT TAAGAATAAA ATCGTGGTGG TGAAAAAAAA 1740  
A 1741

(2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2808 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: OVARTUT03
- (B) CLONE: 2781553

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139 :

GGCAAGATGG CGGAAGGGGA GGACGTGGGA TGGTGGCGGA GCTGGCTGCA GCAGAGCTAC 60  
CAAGCAGTCA AAGAGAAGTC CTCTGAAGCC TTGGAGTTTA TGAAGCGGA CCTGACGGAG 120  
TTTACCCAGG TGGTGCAGCA TGACACGGCC TGTACCATCG CAGCCACGGC CAGCGTGGTC 180  
AAGGAGAAGC TGGCTACGGA AGGCTCCTCA GGAGCAACAG AGAAGATGAA GAAAGGGTTA 240  
TCTGACTTCC TAGGGGTGAT CTCAGACACC TTTGCCCCCT CGCCAGACAA AACCATCGAC 300  
TGCGATGTCA TCACCTGAT GGGCACACCG TCTGGCACAG CTGAGCCCTA TGATGGCACC 360  
AAGGCTCGCC TCTATAGCCT GCAGTCGGAC CCAGCAACCT ACTGTAATGA ACCAGATGGG 420  
CCCCCGGAAT TGTTTGACGC CTGGCTTTCC CAGTTCTGCT TGGAGGAGAA GAAGGGGGAG 480  
ATCTCAGAGC TCCTTGTAGG CAGCCCCCTC ATCCGGGCCC TCTACACCAA GATGGTTCCA 540  
GCAGCTGTTT CCCATTGAGA ATTCTGGCAT CGGTATTTCT ATAAAGTCCA TCAGTTAGAG 600  
CAGGAGCAGG CCCGGAGGGA CGCCCTGAAG CAGCGGGCGG AACAGAGCAT CTCTGAAGAG 660  
CCCGGCTGGG AGGAGGAGGA AGAGGAGCTC ATGGGCATTT CACCCATATC TCCAAAAGAG 720  
GCAAAGGTTT CTGTGGCCAA AATTTCTACA TTCCCTGAAG GAGAACCTGG CCCCAGAGC 780  
CCCTGTGAAG AGAATCTGGT GACTTCAGTT GAGCCCCCAG CAGAGGTGAC TCCATCAGAG 840  
AGCAGTGAGA GCATCTCCCT CGTGACACAG ATCGCCAACC CGGCCACTGC ACCTGAGGCA 900  
CGAGTGCTAC CCAAGGACCT GTCCCAAAAG CTGCTAGAGG CATCCTTGGA GGAACAGGGC 960  
CTGGCTGTGG ATGTGGGTGA GACTGGACCC TCACCCCTA TTAACCTCAA GCCCCTAACG 1020

PF-0459 US

CCTGCTGGCC ACACCGGCGG CCCAGAGCCC AGGCCTCCAG CCAGAGTAGA GACTCTGAGG 1080  
GAGGAGGCGC CCACAGACTT ACGGGTGTTC GAGCTGAAGT CGGATAGTGG GAAGTCTACA 1140  
CCCTCCAACA ATGGAAAGAA AGGCTCAAGC ACGGACATCA GTGAGGACTG GGAGAAAGAC 1200  
TTTGACTTGG ACATGACTGA AGAGGAGGTG CAGATGGCAC TTTCCAAAGT GGATGCCTCC 1260  
GGGGAGCTGG AAGATGTAGA GTGGGAGGAC TGGGAGTGAG GGAGCCAGAG GGAGCAGCTC 1320  
CCCCACCCAT GGCATCTCTC GCCTCCCTCG CTCGTCTCAG CCCAGCCCTG GAAGACTGAG 1380  
AATGTTCCCC CAAATCTCCT CTGCCAACCA GAGCTCTGGG CACAGATTCT GGTGGCTCCC 1440  
TGCTGGCCCT CTTGGGCCTC TGCTCACACC TGGGAAGGGG CTCTCTAAAT CCCGGCCAGA 1500  
AACTCTGACT TGTGCCAACA ATAGGATGAC CCAAGGGAGA GGAAACCTAT CCTCCTCACC 1560  
AGAAGAGCCT GTGTTTTTCT GCTGAACACC CACTGTTTCT GAGGACTCCT GCTGGGAAGT 1620  
CCCAAGGGAT AGTTCTAGCC CTTCTGCCTG TGTAGACAGA AGCTAAACCA CCAGTCTCTC 1680  
TCGGAGGAAG CTGAGACAAC ATACTCTGTC CATAcataag CAGGCAGGGA GGGCCATGCC 1740  
ACCTACCCTT GGCTAAACAG GGACAGTGAA CACATTTTGG TTCCTATCCC AGTGGGTAAG 1800  
AGGCACTTAT CTCTGGGAAA TTTGCCTCTC TTGGGACTCT CCCCCTCCCA GGCATTTTCC 1860  
ATTCTTGGA AGGCTCCTTT GGGGTTTCTG ATCCAGAGAC CAAACCCTGA CCCACCTCCT 1920  
TCCTTTCCTC CAGCCCACGC TGGTCTGTCC CCATGCCTTC CCAGGGCTTC TTCATGTCAG 1980  
ATGCACCCAA GTCCTTAGCC CAGCTGTGCC ACCTGCAGGA GTTCGCTCTT GCGTTTCTTC 2040  
CCCTCCCCAA GAAGGGAGGG GGCTACTTCA GGCCCTTCTG TGTGTTGCCT GGCAGGATAC 2100  
CTTGTTCCAAC CAGCTACCCA CCTCAACTCC CCTGTAGTTT AGGACACAAA ACAGCTACCA 2160  
GCGGTACAGA GCGGTGATCA AAGCCGAGTA CTTACAATC TGGTAAAGCCT AGCTTCTCCG 2220  
CCTCAGCCCT TCTGCTTCTG GAAGGGCTAT CCTGGGGGTG AACTTGAAAC TCTCATCAGG 2280  
CTTCTGCAAA AGCTCTTCTT CCTGAAGACA GACCCAGCCT TTGTGCTCTC ACCCTCCACT 2340  
CTGGTAAAGC TGCACCTCTG GGGGAATGAG GGGCTGCAGG AATCTCTGGA GAGCCTGGTG 2400  
CTTCACGATG CTGCTCTGGT GATTCTTGTA CCTAATCTGG TGTGCTCACC AATGAGTGAA 2460  
AGGGATCGTG GGTGAGGGAC ACCGAGAGAG TGAGGTCACT TCCACTTCAA ACCTTCAGTG 2520  
AGGGGGTGGG ATGGAGAGAA TGCTGAATCT TTTTTTTGAC GGGATGGGGT TTTTCTCTTT 2580  
GTAATTATTT CTTTAGTTTA ATTAACCTTT TGGTTGTTTG TGCAATATTA TATATTTTAA 2640  
ATTATAATGC ATCTCCCCAG AGTATTTTGT AGCTGGGAAA AGAAAAAAGG AAAAAAGAA 2700  
AAAAAGATTC TAACAGCTGT TAGTTTTATA ATTAATAAAG AAAGAAAAAA GAACTTTGTC 2760  
CTGAACCTTT TACAGACTTG CCGTTAACAG CATTAAAGTG ATTCACCC 2808

(2) INFORMATION FOR SEQ ID NO: 140:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 717 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: ADRETUT06  
 (B) CLONE: 2821925

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140 :

CATGCGCCGA CCTTCCTCGG CTGGATTAC ANGTTNNCCC TTAACACCCG GGATTAAAG 60  
 GACCCACACT ACCTTCCCGA AGTTGAAGGC AAGCGGTGAT TGTTTGTAGA CGGCGCTTTG 120  
 TCATGGGACC TGTGCGGTTG GGAATATTGC TTTTCCTTTT TTTGGCCGTG CACGAGGCTT 180  
 GGGCTGGGAT GTTGAAGGAG GAGGACGATG ACACAGAACG CTTGCCCAGC AAATGCGAAG 240  
 TGTGTAAGCT GCTGAGCACA GAGCTACAGG CGGAACTGAG TCGCACCGGT CGATCTCGAG 300  
 AGGTGCTGGA GCTGGGGCAG GTGCTGGATA CAGGCAAGAG GAAGAGACAC GTGCCTTACA 360  
 GCGTTTCAGA GACAAGGCTG GAAGAGGCCT TAGAGAATTT ATGTGAGCGG ATCCTGGACT 420  
 ATAGTGTTCA CGCTGAGCGC AAGGGCTCAC TGAGATATGC CAAGGGTCAG AGTCAGACCA 480  
 TGGCAACACT GAAAGGCCTA GTGCAGAAGG GGGTGAAGGT GGATCTGGGG ATCCCTCTGG 540  
 AGCTTTTGGG ATGAGCCCAG CCGTTGAGGT CACATACCTC AAGAAGCAGT GTGAGACCAT 600  
 GTTNGAGGAG TTTTGAGACA TTGTGGGAGA CTGGTACTTG CACCATCAGG AGCAGCCGCT 660  
 ACAAGATTTT CTCTGTGAAG GTCATGTGCT GCCAGCTGCT TGAAGTGCAT GTCGGGT 717

(2) INFORMATION FOR SEQ ID NO: 141:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2552 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: UTRSTUT05  
 (B) CLONE: 2879068

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141 :

GGCAGGGGGC GCGCCGGGCC CAGCGCCACG TCACCGCCCA GCAGCCCTCC CGATTGGCGG 60  
 GCGGGGCGGC TATAAAGGGA GGGCGCAGGC GGCGCCCGGA TCTCTTCCGC CGCCATTTTA 120



PF-0459 US

AATCCAGCTC CATACAACGC TCCGCCGCCG CTGCTGCCGC GACCCGGACT GCGCGCCAGC 180  
ACCCCCCTGC CGACAGCTCC GTCACTATGG AGGATATGAA CGAGTACAGC AATATAGAGG 240  
AATTCGCAGA GGGATCCAAG ATCAACGCGA GCAAGAATCA GCAGGATGAC GGTAAAATGT 300  
TTATTGGAGG CTTGAGCTGG GATACAAGCA AAAAAGATCT GACAGAGTAC TTGTCTCGAT 360  
TTGGGGAAGT TGTAAGTGC ACAATTAAAA CAGATCCAGT CACTGGGAGA TCAAGAGGAT 420  
TTGGATTTGT GCTTTTCAAA GATGCTGCTA GTGTTGATAA GGTTTTGGAA CTGAAAGAAC 480  
ACAACTGGA TGGCAAATTG ATAGATCCCA AAAGGGCCAA AGCTTTAAAA GGGAAAGAAC 540  
CTCCCAAAAA GGTTTTTGTG GGTGGATTGA GCCCGGATAC TTCTGAAGAA CAAATTAAAG 600  
AATATTTTGG AGCCTTTGGA GAGATTGAAA ATATTGAACT TCCCATGGAT AAAAAACAA 660  
ATGAAAGAAG AGGATTTTGT TTTATCACAT ATACTGATGA AGAGCCAGTA AAAAAATTGT 720  
TAGAAAGCAG ATACCATCAA ATTGGTTCTG GGAAGTGTGA AATCAAAGTT GCACAACCCA 780  
AAGAGGTATA TAGGCAGCAA CAGCAACAAC AAAAAGGTGG AAGAGGTGCT GCAGCTGGTG 840  
GACGAGGTGG TACGAGGGGT CGTGGCCGAG GTCAGGGCCA AACTGGAAC CAAGGATTTA 900  
ATAACTATTA TGATCAAGGA TATGGAAATT ACAATAGTGC CTATGGTGGT GATCAAAACT 960  
ATAGTGGCTA TGGCGGATAT GATTATACTG GGTATAACTA TGGGAACTAT GGATATGGAC 1020  
AGGGATATGC AGACTACAGT GGCCAACAGA GCACTTATGG CAAGGCATCT CGAGGGGGTG 1080  
GCAATCACCA AAACAATTAC CAGCCATACT AAAGGAGAAC ATTGGAGAAA ACAGGTGTGT 1140  
ATAAGAGTAC AGGAAAAACAG TAGAAATGTC TAATTTAATT TAAAGATCAA TAGACAAATG 1200  
AAACGTAAAA ACAAATACT ATGTAGCCTG TTTTACTAA ATTGTTGATT TTTTAATTGC 1260  
TTTATGAGCC TGTTTTGCCT AAAGTGTCTA TAGATCTTTA ACTTTAAAGT CTTATCTCAC 1320  
TTTCTTTAGT ATTGCAGAAA AACTTAAGAG TTTTCTGTT TGCTTTTGTG TACCAGGTGG 1380  
TCTAGAGGAA TAATTAAACA TTTTAGAACT ATTAACAGGT AAAGTACTGA AATGGGTACA 1440  
ACTTAAGGAA AACAAGAATG TTGTCTTCTA ACTCTGACAT TATACCTTGT TTGTACCCGC 1500  
CAGCGGGAAC TTCATTGCAG GCCGTGTGTC ACCCTGACCA CGTCTATCTC TGGGGGTCGC 1560  
ACGTTGCGGG CAGAGCGCAA GGCATACACC AGAAAACGCT GTCCTGTGGT ATGGTCTCTT 1620  
CCAACTTCAT GTACCAGCGT AAAGATTAAA GTGGAAACT TCAGACTTTG GCTTCATTTT 1680  
TAATCTTTTT GGAGATTAAG TGTCTAACT TAACTTAAAT GGTTTTTTAC AGGAGTTAAA 1740  
GTACATAAAT GCCTTTTTAC AGCTTAATCA TTTTGGTCTT CTGTTTAGTG TTGTATTTCA 1800  
ATTGTGGAGC CTCATTTTAA GTGTTTATTC TTTTAAGATT TAATGCTTGC TTTTCTTTT 1860  
TATAGCTAAT AGTGAAATCT ACAAACCAA ACAAGAACTT TTAAATCTGG GATATAAATT 1920  
AAAGATCATA TGCACAGATC AATTTATGTT CTTGTAATAA ACTTATTAGA AATTGGTGTT 1980

PF-0459 US

TGTGATAGCA TTTTACTTGG GTTACTAGAG ATGCTTCTAG TAGACCTTAA TCTAGCATAG 2040  
TTGAACCTCT GAATATGGGA AGGTTGTATT CCCAGATTCT TTCCTGAATA GATTTGAATT 2100  
TAATGTCATT TGGGAACTCC AGGGTGAGTT TATTGACTAC CCAAACCTGTA TTTTACCAAT 2160  
AAATATGCAT ATGATCTTTA ATTATTGAAG AAAATAAAGT GAGGACTTAA AACAATTCAT 2220  
GAAAGTGGAC CTTTAAAGC TTGTCAGAGT TGCACAAATC TAACTGGTAT TTTGTTTTTG 2280  
TTTTTAGGAG GAGATGTTAA AGTAACCCAT CTTGCAGGAC GACATTGAAG ATTGGTCTTC 2340  
TGTTGATCTA AGATGATTAT TTTGTAAAAG ACTTTCTAGT GTACAAGACA CCATTGTGTC 2400  
CAACTGTATA TAGCTGCCAA TTAGTTTTCT TTGTTTTTAC TTTGTCCTTT GCTATCTGTG 2460  
TTATGACTCA ATGTGGATTT GTTTATACAC ATTTTATTTG TATCATTTC A TGTTAAACCT 2520  
CAAATAAATG CTTCCTTATG TGAAAAAAA AA 2552

(2) INFORMATION FOR SEQ ID NO: 142:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1046 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: SINJNOT02  
(B) CLONE: 2886757

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142 :

TACCAGTGTA AAGCCAGAGC TGAGGTTCTT GATAGTCCAC AATGGGTGAA CCACAGCAAG 60  
TGAGTGCACT TCCACCACCT CCAATGCAAT ATATCAAGGA ATATACGGAT GAAAATATTC 120  
AAGAAGGCTT AGCTCCCAAG CCTCCCCCTC CAATAAAAGA CAGTTACATG ATGTTTGGCA 180  
ATCAGTTCCA ATGTGATGAT CTTATCATCC GCCCTTTGGA AAGTCAGGGC ATCGAACGGC 240  
TTCATCCTAT GCAGTTTGAT CACAAGAAAG AACTGAGAAA ACTTAATATG TCTATCCTTA 300  
TTAATTTCTT GGACCTTTTA GATATTTTAA TAAGGAGCCC TGGGAGTATA AAACGAGAAG 360  
AGAAACTAGA AGATCTTAAG CTGCTTTTTG TACACGTGCA TCATCTTATA AATGAATACC 420  
GACCCACCA AGCAAGAGAG ACCTTGAGAG TCATGATGGA GGTCCAGAAA CGTCAACGGC 480  
TTGAAACAGC TGAGAGATTT CAAAAGCACC TGGAACGAGT AATTGAAATG ATTCAGAATT 540  
GCTTGGCTTC TTTGCCTGAT GATTTCCTC ATTGAGAAGC AGGAATGAGA GTAAAACTG 600  
AACCAATGGA TGCTGATGAT AGCAACAATT GTACTGGACA GAATGAACAT CAAAGAGAAA 660  
ATTGAGTCA TAGGAGAGAT CAGATTATAG AGAAAGATGC TGCCTTGTGT GTCCTAATTG 720

PF-0459 US

ATGAGATGAA TGAAAGACCA TGAAAGATGT TTCTTTTCT TTTTTCCTT TTGATAATAG 780  
CATCATATA TAGTTCATTT TCTTTTGGAC AGTCTTAAGA GAAGTTTCAC TAAAAATGTA 840  
AACAGCTTTA ATCTTGACTC CAAATTTTTC AATTATGAGA TGTCATAGGC AGTAATTTTCG 900  
CTGTATAACA AGCATAGACA AATGAGTGTC CCTGCACTAA GAAGAATCAC TTTAAAAAGC 960  
AAAGTGTTAG CTGCTGTTGT ATGGGACATT CCTATGTTTT AGAGTTGCAG TAAACTTTG 1020  
ATGATAACCT CAAAAAATA TAAAAA 1046

(2) INFORMATION FOR SEQ ID NO: 143:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1864 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: SCORNOT04
- (B) CLONE: 2964329

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143 :

GCCCTGGGCT CGCGGCGGTG CCGCGGGGAT GGCGGGAGCC GGAGCTGGAG CCGGAGCTCG 60  
CGGCGGAGCG GCGGCGGGGG TCGAGGCTCG AGCTCGCGAT CCACCGCCCG CGCACCGCGC 120  
ACATCCTCGC CACCCTCGGC CTGCGGCTCA GCCCTCGGCC CGCAGGATGG ATGGCGGGTC 180  
AGGGGGCCTG GGGTCTGGGG ACAACGCCCC GACCACTGAG GCTCTTTTCG TGGCACTGGG 240  
CGCGGGCGTG ACGGCGCTCA GCCATCCCCT GCTCTACGTG AAGCTGCTCA TCCAGGTGGG 300  
TCATGAGCCG ATGCCCCCA CCCTTGGGAC CAATGTGCTG GGGAGGAAGG TCCTCTATCT 360  
GCCGAGCTTC TTCACCTACG CCAAGTACAT CGTGCAAGTG GATGGTAAGA TAGGGCTGTT 420  
CCGAGGCCTG AGTCCCCGGC TGATGTCCAA CGCCCTCTCT ACTGTGACTC GGGGTAGCAT 480  
GAAGAAGGTT TTCCCTCCAG ATGAGATTGA GCAGGTTTCC AACAAGGATG ATATGAAGAC 540  
TTCCCTGAAG AAAGTTGTGA AGGAGACCTC CTACGAGATG ATGATGCAGT GTGTGTCCCG 600  
CATGTTGGCC CACCCCCTGC ATGTCTATCT AATGCGCTGC ATGGTCCAGT TTGTGGGACG 660  
GGAGGCCAAG TACAGTGGTG TGCTGAGCTC CATTGGGAAG ATTTTCAAAG AGGAAGGGCT 720  
GCTGGGATTC TTCGTTGGAT TAATCCCTCA CCTCCTGGGC GATGTGGTTT TCTTGTGGGG 780  
CTGTAACCTG CTGGCCCACT TCATCAATGC CTACCTGGTG GATGACAGCT TCAGCCAGGC 840  
CCTGGCCATC CGGAGCTATA CCAAGTTCGT GATGGGGATT GCAGTGAGCA TGCTGACCTA 900  
CCCCTTCCTG CTAGTTGGCG ACCTCATGGC TGTGAACAAC TGCGGGCTGC AAGCTGGGCT 960

PF-0459 US

CCCCCCTTAC TCCCCAGTGT TCAAATCCTG GATTCACTGC TGGAAGTACC TGAGTGTGCA 1020  
GGGCCAGCTC TTCCGAGGCT CCAGCCTGCT TTTCCGCCGG GTGTCATCAG GATCGTGCTT 1080  
TGCCCTGGAG TAACCTGAAT CATCTAAAAA ACACGGTCTC AACCTGGCCA CCGTGGGTGA 1140  
GGCCTGACCA CCTTGGGACA CCTGCGAGAC GACTCCAACC CAACAACAAC CAGATGTGCT 1200  
CCAGCCCAGC CGGGCTTCAG TTCCATATTT GCCATGTGTC TGTCCAGATG TGGGGTTGAG 1260  
CGGGGGTGGG GCTGCACCCA GTGGATTGGG TCACCCGGCA GACCTAGGGA AGGTGAGGCG 1320  
AGGTGGGGAG TTGGCAGAAT CCCCATACCT CGCAGATTG CTGAGTCTGT CTTGTGCAGA 1380  
GGGCCAGAGA ATGGCTTATG GGGGCCCAGG TTGGATGGGG AAAGGCTAAT GGGGTCAGAC 1440  
CCCACCCCGT CTACCCCTCC AGTCAGCCCA GCGCCCATCC TGCAGCTCAG CTGGGAGCAT 1500  
CATTCTCCTG CTTTGTACAT AGGGTGTGGT CCCCTGGCAC GTGGCCACCA TCATGTCTAG 1560  
GCCTATGCTA GGAGGCAAAT GGCCAGGCTC TGCCTGTGTT TTTCTCAACA CTACTTTTCT 1620  
GATATGAGGG CAGCACCTGC CTCTGAATGG GAAATCATGC AACTACTCAG AATGTGTCCT 1680  
CCTCATCTAA TGCTCATCTG TTTAATGGTG ATGCCTCGCG TACAGGATCT GGTTACCTGT 1740  
GCAGTTGTGA ATACCCAGAG GTTGGGCAGA TCAGTGTCTC TAGTCCTACC CAGTTTTTAA 1800  
GTTTCATGGTA AGATTGACC TCATCTCCCG CAAATAAATG TATTGGTGAT TTGGAACAAA 1860  
AAAA 1864

(2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2295 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: SCORNOT04
- (B) CLONE: 2965248

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144 :

GTCTGCAGCT CCGGCCGCCA CTTGCGCCTC TCCAGCCTCC GCAGGCCCAA CCGCCGCCAG 60  
CACCATGGCC AGCACCATTT CCGCCTACAA GGAGAAGATG AAGGAGCTGT CGGTGCTGTC 120  
GCTCATCTGC TCCTGCTTCT ACACACAGCC GCACCCCAAT ACCGTCTACC AGTACGGGGA 180  
CATGGAGGTG AAGCAGCTGG ACAAGCGGGC CTCAGGCCAG AGCTTCGAGG TCATCCTCAA 240  
GTCCCCTTCT GACCTGTCCC CAGAGAGCCC TATGCTCTCC TCCCCACCCA AGAAGAAGGA 300  
CACCTCCCTG GAGGAGCTGC AAAAGCGGCT GGAGGCAGCC GAGGAGCGGA GGAAGACGCA 360

PF-0459 US

GGAGGCGCAG GTGCTGAAGC AGCTGGCGGA CGGCGCGAGC ACGAGCGCGA GGTGCTGCAC 420  
AAGGCGCTGG AGGAGAATAA CAACTTCAGC CGCCAGGCGG AGGAGAAGCT CAACTACAAG 480  
ATGGAGCTCA GCAAGGAGAT CCGCGAGGCA CACCTGGCCG CACTGCGCGA GCGGCTGCGC 540  
GAGAAGGAGC TGCACGCGGC CGAGGTGCGC AGGAACAAGG AGCAGCGAGA AGAGATGTCG 600  
GGCTAAGGGC CCGGGACGGG CGGCGCCCAT CCTGCGACGG AACACGTTTCG GGTTTTGGTT 660  
TTGTTTCGTT CACCTCTGTC TAGATGCAAC TTTTGTTCCT CCTCCCCAC CCCAGCCCCC 720  
AGCTTCATGC TTCTCTCCG CACTCAGCCG CCCTGCCCTG TCCTCGTGGT GAGTCGCTGA 780  
CCACGGCTTC CCCTGCAGGA GCCGCCGGG GTGAGACGCG GTCCCTCGGT GCAGACACCA 840  
GGCCGGGCGC GGCTGGGTCC CCCGGGGGCC CTGTGAGAGA GGTGGTGGTG ACCGTGGTAA 900  
ACCCAGGGCG GTGGCGTGGG ATCGCGGGTC CTTACGCTGG GCTGTCTGGT CAGCACGTGC 960  
AGGTCAGGGC AGGTCCTCTG AGCCGGCGCC CCTGGCCAGC AGGCGAGGCT ACAGTACCTG 1020  
CTGTCTTTCC AGGGGAAGG GGCTCCCCAT GAGGGAGGGG CGACGGGGGA GGGGGGTGAT 1080  
GGTGCCCTGGG AGCCTGCGTG TGCAGCCGGT GCTTGTGAA CTGGCAGGCG GGTGGGTGGG 1140  
GGCTGCAGCT TTCCTTAATG TGGTTGCACA GGGGTCCTCT GAGACCACCT GGCCTGAGGT 1200  
GGACACCTTG GGCCTTCCTG GAAGCCTGCA GTTGGGGGCC TGCCCTGAGT CTGCTGGGGA 1260  
GTGGGCATTC TCTGCCAGG ACCCATGAGC AGGCTGCATG GTCTAGAGGT TGTGGGCAGC 1320  
ATGGACAGTC CCCCCTCAG AAGTGCAAGA GTTCCAAAGA GCCTCTGGCC CAGGCCCCCTC 1380  
CGTGGGACAG CCCC GCCGCC CCTCCCCACC AGGGCTTTGC AGATGTCCTT GAAAGACCCA 1440  
CCCTAGAGCC CTTTGGAGTG CTGGCCCTC CTGTGCCCTC TGCCCTGGTG GAAGCGGCAG 1500  
CCACAAGTCC TCCTCAGGA GCCCAAGGG GGATTTTGTG GGACCGCTGC CCACAGATCC 1560  
AGGTGTTGGA AGGGCAGCGG GTAAGGTTCC CAAGCCAGCC CCAACACCCT TCCCACTTGG 1620  
CACCCAGAGG GGGCTGTGGG TGGAGGCCTG ACTCCAGGCC TCTCCTGCCC ACACCCTCTG 1680  
GGCTGAGTTC CTTCTTTCCC TTGGACGCC AGTGCTGGCC TTGGAGGACG GTCAGCTGGA 1740  
GGATGGCGGT GGGGGAGGCT GTCTTTGTAC CACTGCAGCA TCCCCACTT CTCCACGGAA 1800  
GCCCCATCCC AAAGCTGCTG CCTGGCCCCT TGCTGTAAAG TGTGAAGGGG GCGGCTGAGT 1860  
TCTCTTAGGA CCCAGAGCCA GGGCCCTCAA CTTCCATCCT GCGGGAGGCC TTGGCCGGGC 1920  
ACTGCCAGTG TCTTCCAGAG CCACACCCAG GGACCACGGG AGGATCCTGA CCCCTGCAGG 1980  
GCTCAGGGGT CAGCAGGGAC CCACTGCCCC ATCTCCCTCT CCCCACCAAG ACAGCCCCAG 2040  
AAGGAGCAGC CAGCTGGGAT GGGAACCCAA GGCTGTCCAC ATCTGGCTTT TGTGGGACTC 2100  
AGAAAGGGAA GCAGAACTGA GGGCTGGGAT ATTCTCATG GTGGCAGCGC TCATAGCGAA 2160  
AGCCTACTGT AATATGCACC CATCTCATCC ACGTAGTAA GTGAACTTAA AAATTCAATC 2220

PF-0459 US

AAATGAACAA TTAAATAAAC ACCTGTGTGT TTAAGACAAA ATAAAAATGG AGGAGAACAA 2280  
AAAAAAAGGG GCGGT 2295

(2) INFORMATION FOR SEQ ID NO: 145:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 842 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: TLYMNOT06
- (B) CLONE: 3000534

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145 :

CGGGGACGGA AGCAGCCCCT GGGCCCGAGG GGCTCGAGGC CGGGCCGGGG CGATGTGGAG 60  
CGCGGGCCGC GGCGGGGCTG CCTGGCCGGT GCTGTTGGGG CTGCTGCTGG CGCTGTTAGT 120  
GCCGGGCGGT GGTGCCGCCA AGACCGGTGC GGAGCTCGTG ACCTGCGGGT CGGTGCTGAA 180  
GCTGCTCAAT ACGCACCACC GCGTGCGGCT GCACTCGCAC GACATCAAAT ACGGATCCGG 240  
CAGCGGCCAG CAATCGGTGA CCGGCGTAGA GCGTCCGAC GACGCCAATA GCTACTGGCG 300  
GATCCGCGGC GGCTCGGAGG GCGGGTGCCC GCGCGGGTCC CCGGTGCGCT GCGGGCAGGC 360  
GGTGAGGCTC ACGCATGTGC TTACGGGCAA GAACCTGCAC ACGCACCCT TCCCGTCGCC 420  
GCTGTCCAAC AACCAGGAGG TGAGTGCCCTT TGGGAAGAC GGCGAGGGCG ACGACCTGGA 480  
CCTATGGACA GTGCGCTGCT CTGGACAGCA CTGGGAGCGT GAGGCTGCTG TGCGCTTCCA 540  
GCATGTGGGC ACCTCTGTGT TCCTGTCAGT CACGGGTGAG CAGTATGGAA GCCCCATCCG 600  
TGGGCAGCAT GAGGTCCACG GCATGCCCAG TGCCAACACG CACAATACGT GGAAGGCCAT 660  
GGAAGGCATC TTCATCAAGC CTAGTGTGGA GCCCTCTGCA GGTACGATG AACTCTGAGT 720  
GTGTGGATGG ATGGGTGGAT GGAGGGTGGC AGGTGGGGCG TCTGCAGGGC CACTCTTGGC 780  
AGAGACTTTG GGTTTGTAGG GGTCTCAAG TGCCTTTGTG ATTAAAGAAT GTTGGTCTAA 840  
AA 842

(2) INFORMATION FOR SEQ ID NO: 146:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2345 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: HEAANOT01  
(B) CLONE: 3046870

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146 :

```

GTCCCGCCCC GCAGCTGCGC GCAGGCGCTC GACGAGCCGC TCGCATTCTA CGTAACGGAC 60
GGCGGAGGCT ACGTGAAGAG AGGCGCGGCG TGA CTGAGCT ACGGTTCTGG CTGCGTCTTA 120
GAGGCATCCG GGGCAGTAAA ACCGCTGCGA TCGCGGAGGC GGCGGCCAGG CCGAGAGGCA 180
GGCCGGGCAG GGGTGTCTGA CGCAGGGCGC TGGGCCGGGT TTCGGCTTCG GCCACAGCTT 240
TTTTTCTCAA GGTGCAATGA AAGCCTTCCA CACTTTCTGT GTTGTCTTC TGGTGTCTTG 300
GAGTGTCTCT GAAGCCAAGT TTGATGATTT TGAGGATGAG GAGGACATAG TAGAGTATGA 360
TGATAATGAC TTCGCTGAAT TTGAGGATGT CATGGAAGAC TCTGTTACTG AATCTCCTCA 420
ACGGGTCATA ATCACTGAAG ATGATGAAGA TGAGACCACT GTGGAGTTGG AAGGGCAGGA 480
TGAAAACCAA GAAGGAGATT TTGAAGATGC AGATACCCAG GAGGGAGATA CTGAGAGTGA 540
ACCATATGAT GATGAAGAAT TTGAAGGTTA TGAAGACAAA CCAGATACTT CTTCTAGCAA 600
AAATAAAGAC CCAATAACGA TTGTTGATGT TCCTGCACAC CTCCAGAACA GCTGGGAGAG 660
TTATTATCTA GAAATTTTGA TGGTGACTGG TCTGCTTGCT TATATCATGA ATTACATCAT 720
TGGAAGAAT AAAACAGTC GCCTTGCACA GGCCTGGTTT AACACTCATA GGGAGCTTTT 780
GGAGAGCAAC TTTACTTTAG TGGGGGATGA TGGAAC TAAC AAGCCA CAAGCACAGG 840
AAAGTTGAAC CAGGAGAATG AGCACATCTA TAACCTGTGG TGTTCTGGTC GAGTGTGCTG 900
TGAGGGCATG CTTATCCAGC TGAGGTTCCCT CAAGAGACAA GACTTACTGA ATGTCTGGC 960
CCGGATGATG AGGCCAGTGA GTGATCAAGT GCAAATAAAA GTAACCATGA ATGATGAAGA 1020
CATGGATACC TACGTATTTG CTGTTGGCAC ACGGAAAGCC TTGGTGCGAC TACAGAAAGA 1080
GATGCAGGAT TTGAGTGAGT TTTGTAGTGA TAAACCTAAG TCTGGAGCAA AGTATGGACT 1140
GCCGACTCT TTGGCCATCC TGTCAGAGAT GGGAGAAGTC ACAGACGGAA TGATGGATAC 1200
AAAGATGGTT CACTTTCTTA CACACTATGC TGACAAGATT GAATCTGTTT ATTTTTCAGA 1260
CCAGTTCTCT GGTCCAAAAA TTATGCAAGA GGAAGGTCAG CCTTTAAAGC TACCTGACAC 1320
TAAGAGGACA CTGTTGTTTA CATTTAATGT GCCTGGCTCA GGTAACACTT ACCCAAAGGA 1380
TATGGAGGCA CTGCTACCCC TGATGAACAT GGTGATTTAT TCTATTGATA AAGCCAAAAA 1440
GTTCCGACTC AACAGAGAAG GCAAACAAAA AGCAGATAAG AACCGTGCCC GAGTAGAAGA 1500
GAACTTCTTG AAAGTACAC ATGTGCAAAG ACAGGAAGCA GCACAGTCTC GGCGGGAGGA 1560

```

PF-0459 US

GAAGAGGAGG AGCGAATCAT GAATGAGGAA GATCCTGAGA AACAGCGCAG 1620  
GCTGGAGGAG GCTGCATTGA GGCCTGAGCA AAAGAAGTTG GAAAAGAAGC AAATGAAAAT 1680  
GAAACAAATC AAAGTGAAAG CCATGTAAAG CCATCCCAGA GATTTGAGTT CTGATGCCAC 1740  
CTGTAAGCTC TGAATTCACA GGAAACATGA AAAACGCCAG TCCATTTCTC AACCTTAAAT 1800  
TTCAGACAGT CTTGGGCAAC TGAGAAATCC TTATTTTCATC ATCTACTCTG TTTGGGGTTT 1860  
GGGGTTTTAC AGAGATTGAA GATACCTGGA AAGGGCTCTG TTTCAAGAAT TTTTTTTTCC 1920  
AGATAATCAA ATTATTTTGA TTATTTTATA AAAGGAATGA TCTATGAAAT CTGTGTAGGT 1980  
TTTAAATATT TTAATAATTA TAATACAAAT CATCAGTGCT TTTAGTACTT CAGTGTTTAA 2040  
AGAAATACCA TGAAATTTAT AGGTAGATAA CCAGATTGTT GCTTTTTGTT TAAACCAAGC 2100  
AGTTGAAATG GCTATAAAGA CTGACTCTAA ACCAAGATTC TGCAAATAAT GATTGGAATT 2160  
GCACAATAAA CATTGCTTGA TGTTTTCTTG TATGTCTACA TTAAACTTGA GAAAAAGTAA 2220  
AAATTAGAAC ACTGTATGTA GTAATGAAAT TTCAGGGACC CAGAACATAA TGTAGTATAT 2280  
GTTTTTAGGT GGGAGATGCT GATAACAAA TTAATAGGAA GTCTGTAGGC ATTAGGATAC 2340  
TGACA 2345

(2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2215 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: PONSAT01
- (B) CLONE: 3057669

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147 :

CCCACGCGTC CGCCCACGCG TCCGTTTTCA GTAGGGATTT CCTGTGACCA GACAAGTTCA 60  
TCTGAGAGCC AGTTCTCACC ACTGGAATTC TCAGGAATGG ACCATGAGGA CATCAGTGAG 120  
TCAGTGGATG CAGCATACAA CCTCCAGGAC AGTTGCCTTA CAGACTGTGA TGTGGAAGAT 180  
GGGACTATGG ATGGCAATGA TGAGGGGCAC TCCTTTGAAC TTTGTCCTTC TGAAGCTTCT 240  
CCTTATGTAA GGTCAAGGGA GAGAACCTCC TCTTCAATAG TATTTGAAGA TTCTGGCTGT 300  
GATAATGCTT CCAGTAAAGA AGAGCCGAAA ACTAATCGAT TGCATATTGG CAACCATTGT 360  
GCTAATAAAC TAACTGCTTT CAAGCCCACC AGTAGCAAAT CTTCTTCTGA AGCTACATTG 420  
TCTATTTCTC CTCCAAGACC AACCCTTTA AGTTTAGATC TCACTAAAAA CACCACAGAA 480



PF-0459 US

AAACTCCAGC CCAGTTCACC AAAGGTGTAT CTTTACATTC AAATGCAGCT GTGCAGAAAA 540  
GAAAACCTCA AAGACTGGAT GAATGGACGA TGTACCATAG AGGAGAGAGA GAGGAGCGTG 600  
TGTCTGCACA TCTTCCTGCA GATCGCAGAG GCAGTGGAGT TTCTTCACAG TAAAGGACTG 660  
ATGCACAGGG ACCTCAAGCC ATCCAACATA TTCTTTACAA TGGATGATGT GGTCAAGGTT 720  
GGAGACTTTG GGTTAGTGAC TGCAATGGAC CAGGATGAGG AAGAGCAGAC GGTTCGACC 780  
CCAATGCCAG CTTATGCCAG ACACACAGGA CAAGTAGGGA CCAAACCTGTA TATGAGCCCA 840  
GAGCAGATTC ATGGAAACAG CTATTCTCAT AAAGTGGACA TCTTTTCTTT AGGCCTGATT 900  
CTATTTGAAT TGCTGTATCC ATTCAGCACT CAGATGGAGA GAGTCAGGAC CTAACTGAT 960  
GTAAGAAATC TCAAATTTCC ACCATTATTT ACTCAGAAAT ATCCTTGTGA GTACGTGATG 1020  
GTTCAAGACA TGCTCTCTCC ATCCCCCATG GAACGACCTG AAGCTATAAA CATCATTGAA 1080  
AATGCTGTAT TTGAGGACTT GGACTTTCCA GGAAAAACAG TGCTCAGACA GAGGTCTCGC 1140  
TCCTTGAGTT CATCGGGAAC AAAACATTCA AGACAGTCCA ACAACTCCCA TAGCCCTTTG 1200  
CCAAGCAATT AGCCTTAAGT TGTGCTAGCA ACCCTAATAG GTGATGCAGA TAATAGCCTA 1260  
CTTCTTAGAA TATGCCTGTC CAAAATTGCA GACTTGAAAA GTTTGTTCTT CGCTCAATTT 1320  
TTTTGTGGAC TACTTTTTTT ATATCAAATT TAAGCTGGAT TTGGGGGCAT AACCTAATTT 1380  
GAGCCAACTC CTGAGTTTTG CTATACTTAA GGAAAGGGCT ATCTTTGTTC TTTGTTAGTC 1440  
TCTTGAAACT GGCTGCTGGC CAAGCTTTAT AGCCCTCACC ATTTGCCTAA GGAGGTAGCA 1500  
GCAATCCCTA ATATATATAT ATAGTGAGAA CTAAAATGGA TATATTTTTA TAATGCAGAA 1560  
GAAGGAAAGT CCCCCTGTGT GGTAAGTGTG TTGTTCTAGA AATATGCTTT CTAGAGATAT 1620  
GATGATTTTG AAAGTGAATT CTAGAAAAAG CTGACTCCAT TTTTGTCCCT GGCGGGTAAA 1680  
TTAGGAATCT GCACTATTTT GGAGGACAAG TAGCACAAAC TGTATAACGG TTTATGTCCG 1740  
TAGTTTTATA GTCCTATTTG TAGCATTCAA TAGCTTTATT CCTTAGATGG TTCTAGGGTG 1800  
GGTTTACAGC TTTTGTACT TTTACCTCCA ATAAAGGGAA AATGAAGCTT TTTATGTAAA 1860  
TTGGTTGAAA GGTCTAGTTT TGGGAGGAAA AAAGCCGTAG TAAGAAATGG ATCATATATA 1920  
TTACAACATA CTTCTTCAAC TATGGACTTT TTAAGCCTAA TGAAATCTTA AGTGTCTTAT 1980  
ATGTAATCCT GTAGGTTGGT ACTTCCCCCA AACTGATTAT AGGTAACAGT TTAATCATCT 2040  
CACTTGCTAA CATGTTTTTA TTTTCACTG TAAATATGTT TATGTTTTAT TTATAAAAAT 2100  
TCTGAAATCA ATCCATTTGG GTTGGTGGTG TACAGAACAC ACTTAAGTGT GTTAAGTTGT 2160  
GACTTCTTTC AAGTCTAAAT GATTTAATAA AACTTTTTTT AAATTAAAAA AAAAA 2215

(2) INFORMATION FOR SEQ ID NO: 148:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1395 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: HEAONOT03
- (B) CLONE: 3088178

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148 :

```

GGTTGACATG ATGAACAATC GGTTCGGAA GGATATGATG AAAAATGCTA GTGAAAGTAA 60
ACTTTCGAAA GACAACCTTA AAAAGAGACT TAAAGAAGAA TTCCAACATG CCATGGGAGG 120
AGTACCTGCC TGGGCAGAGA CTACTAAGCG GAAAACATCT TCAGATGATG AAAGTGAAGA 180
GGATGAAGAT GATTTGTTGC AAAGGACTGG GAATTCATA TCCACATCAA CTTCTCTTCC 240
AAGAGGCATC TTGAAGATGA AGAACTGCCA GCATGCGAAT GCTGAACGTC CTACTGTTGC 300
TCGGATCTCA TCTGTGCAGT TCCATCCCGG TGCACAGATT GTGATGGTTG CTGGATTAGA 360
TAATGCTGTA TCACTATTTT AGGTTGATGG GAAAACAAAT CCTAAAATTC AGAGCATCTA 420
TTTGAAAGG TTTCCAATCT TTAAGGCTTG TTTTAGTGCT AATGGGGAAG AAGTTTTAGC 480
CACGAGTACC CACAGCAAGG TTCTTTATGT CTATGACATG CTGGCTGGAA AGTTAATTCC 540
TGTGCATCAA GTGAGAGGTT TGAAAGAGAA GATAGTGAGG AGCTTTGAAG TCTCCCCAGA 600
TGGGTCCTTC TTGCTCATAA ATGGCATTGC TGGATATTTG CATTTGCTAG CAATGAAGAC 660
CAAAGAACTG ATTGGAAGCA TGAAAATTAA TGGAAGGGTT GCAGCATCCA CATTCTCTTC 720
AGATAGTAAG AAAGTATACG CCTCTTCGGG GGATGGAGAA GTTTATGTTT GGGATGTGAA 780
CTCAAGGAAG TGCCTTAACA GATTTGTTGA TGAAGGCAGT TTATATGGAT TAAGCATTGC 840
CACATCTAGG AATGGACAGT ATGTTGCTTG TGGTTCTAAT TGTGGAGTGG TAAATATATA 900
CAATCAAGAT TCTTGTCTCC AAGAAACAAA CCCAAAGCCA ATAAAAGCTA TAATGAACTT 960
GGTTACAGGT GTTACTTCTC TGACCTTCAA TCCTACTACA GAAATCTTGG CAATTGCTTC 1020
AGAAAAATG AAAGAAGCAG TCAGATTGGT TCATCTTCCT TCCTGTACAG TATTTTCAAA 1080
CTTCCCAGTC ATTAAAAATA AGAATATTTT TCATGTTTAT ACCATGGATT TTTCTCCGAG 1140
AAGTGGATAC TTTGCCTTGG GGAATGAAAA GGGCAAGGCC CTGATGTATA GGTTGCACCA 1200
TTACTCAGAC TTCTAAAGAG ACTATTTGAA GTCCAGTTGA GTCACAAGAG AAGCCTGTCT 1260
TGATATATCA TCTCAGAAAC TTTCTGAAT ATGTGATAAT ATATGGAAAA TGATTTATAG 1320
ATCCAGCTGT GCTTAAGAGC CAGTAATGTC TTAATAAACA TGTGGCAGCT TTTGTTTGAA 1380
AAAAAAAAAA AAAGG

```

1395

(2) INFORMATION FOR SEQ ID NO: 149:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2609 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRSTNOT19
- (B) CLONE: 3094321

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149 :

```

CCCCCCATGG CACTGTCGCG GGGGCTGCCC CGGGAGCTGG CTGAGGCGGT GGCCGGGGGC 60
CGGGTGCTGG TGGTGGGGGC GGGCGGCATC GGCTGCGAGC TCCTCAAGAA TCTCGTGCTC 120
ACCGGTTTCT CCCACATCGA CCTGATTGAT CTGGATACTA TTGATGTAAG CAACCTCAAC 180
AGACAGTTTT TGTTTCAAAA GAAACATGTT GGAAGATCAA AGGCACAGGT TGCCAAGGAA 240
AGTGTACTGC AGTTTTACCC GAAAGCTAAT ATCGTTGCCT ACCATGACAG CATCATGAAC 300
CCTGACTATA ATGTGGAATT TTTCCGACAG TTTATACTGG TTATGAATGC TTTAGATAAC 360
AGAGCTGCCC GAAACCATGT TAATAGAATG TGCCTGGCAG CTGATGTTCC TCTTATTGAA 420
AGTGGAACAG CTGGGTATCT TGGACAAGTA ACTACTATCA AAAAGGGTGT GACCGAGTGT 480
TATGAGTGTC ATCCTAAGCC GACCCAGAGA ACCTTTCCTG GCTGTACAAT TCGTAACACA 540
CCTTCAGAAC CTATACATTG CATCGTTTGG GCAAAGTACT TGTTCAACCA GTTGTGTTGGG 600
GAAGAAGATG CTGATCAAGA AGTATCTCCT GACAGAGCTG ACCCTGAAGC TGCCTGGGAA 660
CCAACGGAAG CCGAAGCCAG AGCTAGAGCA TCTAATGAAG ATGGTGACAT TAAACGTATT 720
TCTACTAAGG AATGGGCTAA ATCAACTGGA TATGATCCAG TTAAACTTTT TACCAAGCTT 780
TTTAAAGATG ACATCAGGTA TCTGTTGACA ATGGACAAAC TATGGCGGAA AAGGAAACCT 840
CCAGTTCCTG TGGACTGGGC TGAAGTACAA AGTCAAGGAG AAGAAACGAA TGCATCAGAT 900
CAACAGAATG AACCCAGTT AGGCCTGAAA GACCAGCAGG TTCTAGATGT AAAGAGCTAT 960
GCACGTCTTT TTTCAAAGAG CATCGAGACT TTGAGAGTTC ATTTAGCAGA AAAGGGGGAT 1020
GGAGCTGAGC TCATATGGGA TAAGGATGAC CCATCTGCAA TGGATTTTGT CACCTCTGCT 1080
GCAAACCTCA GGATGCATAT TTTCACTATG AATATGAAGA GTAGATTTGA TATCAAATCA 1140
ATGGCAGGGA ACATTATTCC TGCTATTGCT ACTACTAATG CAGTAATTGC TGGGTTGATA 1200
GTATTGGAAG GATTGAAGAT TTTATCAGGA AAAATAGACC AGTGCAGAAC AATTTTTTTG 1260
AATAACAAC CAAACCCAAG AAAGAAGCTT CTTGTGCCTT GTGCACTGGA TCCTCCCAAC 1320

```

PF-0459 US

CCCAATTGTT ATGTATGTGC CAGCAAGCCA GAGGTGACTG TGCGGCTGAA TGTCCATAAA 1380  
GTGACTGTTC TCACCTTACA AGACAAGATA GTGAAAAGAAA AATTTGCTAT GGTAGCACCA 1440  
GATGTCCAAA TTGAAGATGG GAAAGGAACA ATCCTAATAT CTTCCGAAGA GGGAGAGACG 1500  
GAAGCTAATA ATCACAAGAA GTTGTCTAGAA TTTGGAATTA GAAATGGCAG CCGGCTTCAA 1560  
GCAGATGACT TCCTCCAGGA CTATACTTTA TTGATCAACA TCCTTCATAG TGAAGACCTA 1620  
GGAAAGGACG TTGAATTTGA AGTTGTTGGT GATGCCCCGG AAAAAGTGGG GCCCAAACAA 1680  
GCTGAAGATG CTGCCAAAAG CATAACCAAT GGCAGTGATG ATGGAGCTCA GCCCTCCACC 1740  
TCCACAGCTC AAGAGCAAGA TGACGTTCTC ATAGTTGATT CGGATGAAGA AGATTCTTCA 1800  
AATAATGCCG ACGTCAGTGA AGAAGAGAGA AGCCGCAAGA GGAAATTAGA TGAGAAAGAG 1860  
AATCTCAGTG CAAAGAGGTC ACGTATAGAA CAGAAGGAAG AGCTTGATGA TGTCATAGCA 1920  
TTAGATTGAA CAGAAATGCC TCTAAACAGA ACCCTCTTAC TATTTAGTTT ATCTGGGCAG 1980  
AACCAGATTG TTATGTCCTT TGTTCCAAAG GGAAAAAATT GACAGCAGTG ACTTGAAAAT 2040  
GATTCTGCTC CCTTTGAAAG CATTCATTTT GCTAGAACTG TTAGACACAT TGCAGTATGC 2100  
TGTATTGAAA GTAGGAATAT AGTTTTAAAA ACCCTTTGAA CAAAGTGTGT GCATAACCAG 2160  
TCATGAGATA AAACAACACA ATGCATGTTG CCTTTTAAAT GTAAATACCC TTAGGTATCA 2220  
TTAATAGTTT CAAAATATTG TGGTTTAGTA AAGTTGATAC CTGGTTATAA ATATTATGCC 2280  
TTTTATTTTG GCTAGAAGAA GAATTATTTT TAGCCTAGAT CTAACCATTT TCATACTCTT 2340  
AACTGATTGA AACAGATTCA AAGAAGTATC GAGTGCTATG CATTGAAACT TGTTTTTAAA 2400  
TGTTAGATGG CACTATGTAT ATTAATGTAA AACAATGTTA ATTTACTCAA GTTTTCAGTT 2460  
TGTACGCCT GGTATGTCTG TGTAAGAAGC CAATTTTGT GTATTGTTAC AGTTTCAGGT 2520  
TATTTATATT CGATGTTTTG TAAACTCAA ATAACGACTA TACTTATGGA CCAAATAAAT 2580  
GGCATCTGCA TTCTTGTTAA AAAAAAAAAA 2609

(2) INFORMATION FOR SEQ ID NO: 150:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3633 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGTUT13
- (B) CLONE: 3115936

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150 :

PF-0459 US

CCTGAGGGAT CCACAGAGGG TGCGGTCCTT GGAGGGAGGA CATGCAGTGC CACGTGCCAT 60  
GGACCAGCCA GTGGACCCCA TGGCCAGCAA GGCTGCTCCT GGGGCCAGTG GGGTGGACAG 120  
TCCCGCCCAC GCAGGTGACT GAGGTGCCAG TGTGGGAATG AAAATGCGGC CTGTGCTCCT 180  
GGGCCCATGC GTCTCACGCT GCCCTTCCTC TCCAGGGAAG CCTGTGTACC TGCTACTTTT 240  
TCCCGAACAA TTCATGGTAA AAACACAAAT GGTATATGGA CAAGATACTG AATGTGGAAG 300  
AAACCTACTT GACAGTGTG GTGAAAATAG GGCCAGGATT TCACACCCGT GAATGCTTTT 360  
TACTGAAAAG TATTTTGTGT TTTTCTCCCA GTTACAGAAT GTCTGAAGGG GACAGTGTGG 420  
GAGAATCCGT CCATGGGAAA CCTTCGGTGG TGTACAGATT TTTACAAGA CTTGGACAGA 480  
TTTATCAGTC CTGGCTAGAC AAGTCCACAC CCTACACGGC TGTGCGATGG GTCGTGACAC 540  
TGGGCCTGAG CTTTGTCTAC ATGATTCGAG TTTACCTGCT GCAGGGTTGG TACATTGTGA 600  
CCTATGCCTT GGGGATCTAC CATCTAAATC TTTTCATAGC TTTTCTTTCT CCCAAAGTGG 660  
ATCCTTCCTT AATGGAAGAC TCAGATGACG GTCCTTCGCT ACCCACCAAA CAGAACGAGG 720  
AATTCCGCCC CTTCAATCGA AGGCTCCCAG AGTTTAAATT TTGGCATGCG GCTACCAAGG 780  
GCATCCTTGT GGCTATGGTC TGTACTTTCT TCGACGCTTT CAACGTCCCG GTGTTCTGGC 840  
CGATTCTGGT GATGTACTTC ATCATGCTCT TCTGTATCAC GATGAAGAGG CAAATCAAGC 900  
ACATGATTAA GTACCGGTAC ATCCCGTTCA CACATGGGAA GAGAAGGTAC AGAGGCAAGG 960  
AGGATGCCGG CAAGGCCTTC GCCAGCTAGA AGCGGGACTG AGGCTGCCTC ACGTGTTGCA 1020  
AGAACAGTTT TGAGCCATTG TTAACAATGC CTTTTTTCTT CACATAAAGT AGTTGATTAC 1080  
GAGGGAGTCA AATTTTCTTT TTA AAAAGGA GCTTCAATGA TTTGTAAGT AAATATCAGG 1140  
TTCTAGAAGA AACTGGCGCT TAAACCAAAT CGCATGGATT TCTTTTTCAG TGACGTTCAA 1200  
GTGTTTCTCA CGGATGGAAT TCTAGTCAGC TGCAGGCGGG AAGCCAGGCG GGTGGAGCCC 1260  
ATGGGAGCAA GGGCGAGTGG CCGGTCCCCG CTGTGCCAGG TGGGCAGGCA GGAGCAAGGC 1320  
CTGCGAGGGA GGAACGGGCC GCTCCCCGCC AGCCGCCTTC CCCAGCAGCC GCAGGTGGTG 1380  
CCAGCCACTC CACAGAGCCC GAGGGATGAT CTAGCCTGAT TCCTGCGTGT CCGAAAGAAC 1440  
TTAACGTTTT AAAGGTGATT GTCAAGTAAC TGTGTGGGGT TCTAATGCCA GTTTCCTAAT 1500  
TCCATCTCAC TGGAGATGTT TAAAGTTGGC CTCTATCCTA ATGACTCAAA ACTTGTTTCT 1560  
TAACTACCAT GATTGCTTTT GAGGGCCCGG AATTATAAAT ATATATTATA TTTTAATTGT 1620  
TTGAGATTAT TTTGACACAT TTCTTTGATA CGTAGAGTGT TTTGTTTTTA ATTTAAATCT 1680  
GTCCTCATGC AACCTCCAT GAGGGGCAGC GAAGCTGGCA GGGAGCAGAC TGGCTTTGTA 1740  
GGTTCAGCAC TCGGCCCCC ACTGCGGGAG AGGCGGAACC CACTTGCATG TCAGCGTTTT 1800  
TGATTGAGA AAAGAAATAC TCTCAACGTT TTACCAAGTG ATTTTACCTC CACCTTTACT 1860

PF-0459 US

AAAGTCTTTA CCTAAAACAT GGCAGTCGCT GGACACAGGA AAGCCCACCT TTTGTTTGGC 1920  
CTTTTCGAAA GGTGACCCAT ATTGCACAGC AGAACATCAC AGCTGTGGTC CCAGATGAGA 1980  
CACTGACATG CGAGTGAAGG CCTCTCCTCC TGGGCCCCGG GCTGCGCAGG CTCCTCACTC 2040  
TGGGCGGTGT TTCCTGTCTC AGAATTGACA CGGTGAATGC TTAGTGTCTG GATTTTCTTG 2100  
TGCCAGTGTT TACATATCTG ACATCGAGCT CCTCTAAGAG GCCACGTTCA AGCTTGTGTG 2160  
TCCCTGACCC AAGATAGCCA GTGCTGCTCC CAGGTGGTAC TTCTGGTACC GTGTTGAGAC 2220  
ACTTGGGATT CTCAGACTGT GGACAGGAGT GTTTGTCATT TTTCATACTG TTTTCTTAAT 2280  
AAGCGCTCAG GCCTAAGGTG TGACAGGAAG TCGCACGCGC TTGGCCAGAG CACAGTGAAG 2340  
CAAAGGACTG GGTGCTGATG GATGGAGCCA CGGCGGCATC TGCCCACCCG GCCGCAGCCC 2400  
CCAGTGCCTC TCCTGGTGGT CCTCCCAGTC TAGAGGGTCA CGGCCCCCCC GCCCTCCTCC 2460  
GTCTCTGGCA AGCTGACCTT GACTAACCCA GGAATACAGG GTCATCCTCA TTCCTAAGTA 2520  
AGTCAAACAG CAAGACATGG TTTGCGCGGG TCTTTGCCGG AAGCCGGTCC TGCTGGCCAG 2580  
GTGTTTTACG TCAGCAGGGA AATGTGGCAC ACGCCCTCGA GGCATTTTAA CACTGTGCTT 2640  
CAGGAAATCT CAAGTTCCAT CTTGTGTAG TAACGTACCC ACATTTTGCT GGAGTTAGTT 2700  
TATTAAAGAT GCCTACGGTG AACTCTCTGG CGCAGGTAA ATGCAGTTTT GAAAACCTGG 2760  
AAACATCAAA TGGAGGCGGG AAATAGGCTG GGGCCGAGCT GAGGGGCTGA ACACAGCAGT 2820  
GACCGTGGGT CAGCAGGTCG CCTGCCCAGC AGGCCCCCA GGAGAGGGCT CGGGCGCCCC 2880  
TGGCAGCCCC CATACCCCCA GGACCTGGCT CGTGAGTGCG TCTGGGTCAG GAAGAGACCT 2940  
CTCTGTGCGT CTCAGGCTGA GATGCAGATT TCTGTTTTCT AAAACTGGAA GCGACCTTGA 3000  
CGTGTATTGA AGGTGTGTGT GCCAAATGCT TCCGACGGAG GTGCTGGCCT TGTTTGTTTT 3060  
CTCTCTGCCC CGTGTGGTCA TCAAGTCCTG GGGGATGTGC TCTGCCCAGC CGCCCTCGGG 3120  
GAGAGCAGCG CCGCCTCCCA TGGGGCCGTG GGGCTGCTGT TCTCACTGCA CTGGCTGAAG 3180  
CAACCCGCCA GCCTCCGTGC CCCACCCAC CCAGCACGCA CTCATTGAGT CCATTGCCTT 3240  
AACACAAGCC TGATGGGGCT GTTTTCTCAC AATATAAAGC AATAAAGTGT CTTCTGGCCT 3300  
ACTTCTGAAT TACTTCTCAA CTGTATGGTT TGGGGAAGGG AGGGAAACCT AAAATCCCGT 3360  
CCAAATAAGT GAAATTCCTG AAGAAGTGGC TGAGTCCTAC CAGGTGGGG TTAGGGAAAT 3420  
GTTCTGGGTT CAGGCGCCCC TCCCAGGGCT GAGAAAGCGC AGCCAGGGAC AGCTTTCTGT 3480  
TCTCTCCCAG GGTGGCTAGG TTAGTATCTT ACATGACAAA AAACCTGAGAG TGTTCTAACT 3540  
TCTGTGCAAG CAAGGTTAAT CCTGAGACTA AATCTTGGCG TTCAGACTCC CGTAGAGGTC 3600  
ATCTGTGTCC AGGCCACCC GGGCGCCGGC TCA 3633

(2) INFORMATION FOR SEQ ID NO: 151:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2018 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
 (A) LIBRARY: LUNGTUT13  
 (B) CLONE: 3116522

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151 :

TGGCTCGCTG GCCGCTCCTG GAGGCGGCGG CGGGAGCGCA GGGGGCGCGC GGCCCGGGGA 60  
 CTCGCATTCC CCGGTTCCCC CTCCACCCCA CGCGGCCTGG ACCATGGACG CCAGATGGTG 120  
 GGCAGTGGTG GTGCTGGCTG CGTTCCCCTC CCTAGGGGCA GGTGGGGAGA CTCCCGAAGC 180  
 CCCTCCGGAG TCATGGACCC AGCTATGGTT CTTCCGATTT GTGGTGAATG CTGCTGGCTA 240  
 TGCCAGCTTT ATGGTACCTG GCTACCTCCT GGTGCAGTAC TTCAGGCGGA AGAACTACCT 300  
 GGAGACCGGT AGGGGCGCTCT GCTTTCCCCT GGTGAAAGCT TGTGTGTTTG GCAATGAGCC 360  
 CAAGGCCTCT GATGAGGTTT CCCTGGCGCC CCGAACAGAG GCGGCAGAGA CCACCCCGAT 420  
 GTGGCAGGCC CTGAAGCTGC TCTTCTGTGC CACAGGGCTC CAGGTGTCTT ATCTGACTTG 480  
 GGGTGTGCTG CAGGAAAGAG TGATGACCCG CAGCTATGGG GCCACAGCCA CATCACC GGG 540  
 TGAGCGCTTT ACGGACTCGC AGTTCCTGGT GCTAATGAAC CGAGTGCTGG CACTGATTGT 600  
 GGCTGGCCTC TCCTGTGTTT TCTGCAAGCA GCCCCGGCAT GGGGCACCCA TGTACCGGTA 660  
 CTCCTTTGCC AGCCTGTCCA ATGTGCTTAG CAGCTGGTGC CAATACGAAG CTCTTAAGTT 720  
 CGTCAGCTTC CCCACCCAGG TGCTGGCCAA GGCCTCTAAG GTGATCCCTG TCATGCTGAT 780  
 GGGAAAGCTT GTGTCTCGGC GCAGCTACGA AACTGGGAG TACCTGACAG CCACCCTCAT 840  
 CTCCATTGGG GTCAGCATGT TTCTGCTATC CAGCGGACCA GAGCCCCGCA GCTCCCCAGC 900  
 CACCACACTC TCAGGCCTCA TCTTACTGGC AGGTTATATT GCTTTTGACA GCTTCACCTC 960  
 AACTGGCAG GATGCCCTGT TTGCCTATAA GATGTCATCG GTGCAGATGA TGTTTGGGGT 1020  
 CAATTTCTTC TCCTGCCTCT TCACAGTGGG CTCACTGCTA GAACAGGGGG CCCTACTGGA 1080  
 GGGAACCCGC TTCATGGGGC GACACAGTGA GTTTGCTGCC CATGCCCTGC TACTCTCCAT 1140  
 CTGCTCCGCA TGTGGCCAGC TCTTCATCTT TTACACCATT GGGCAGTTTG GGGCTGCCGT 1200  
 CTTACCATC ATCATGACCC TCCGCCAGGC CTTTGCCATC CTTCTTTCTT GCCTTCTCTA 1260  
 TGGCCACACT GTCAGTGTGG TGGGAGGGCT GGGGGTGGCT GTGGTCTTTG CTGCCCTCCT 1320  
 GCTCAGAGTC TACGCGCGGG GCCGTCTAAA GCAACGGGGA AAGAAGGCTG TGCCTGTTGA 1380

PF-0459 US

GTCTCCTGTG CAGAAGGTTT GAGGGTGGAA AGGGCCTGAG GGGTGAAGTG AAATAGGACC 1440  
CTCCCACCAT CCCCTTCTGC TGTAACTCT GAGGGAGCTG GCTGAAAGGG CAAAATGCAG 1500  
GTGTTTTCTC AGTATCACAG ACCAGCTCTG CAGCAGGGGA TTGGGGAGCC CAGGAGGCAG 1560  
CCTTCCCTTT TGCCTTAAGT CACCCATCTT CCAGTAAGCA GTTTATTCTG AGCCCCGGGG 1620  
GTAGACAGTC CTCAGTGAGG GGTTTTGGGG AGTTTGGGGT CAAGAGAGCA TAGGTAGGTT 1680  
CCACAGTTAC TCTTCCCACA AGTTCCTTA AGTCTTGCCC TAGCTGTGCT CTGCCACCTT 1740  
CCAGACTCAC TCCCCTCTGC AAATACCTGC ATTTCTTACC CTGGTGAGAA AAGCACAAGC 1800  
GGTGTAGGCT CCAATGCTGC TTTCCCAGGA GGGTGAAGAT GGTGCTGTGC TGAGGAAAGG 1860  
GGATGCAGAG CCCTGCCCAG CACCACCACC TCCTATGCTC CTGGATCCCT AGGCTCTGTT 1920  
CCATGAGCCT GTTGCAGGTT TTGGTACTTT AGAAATGTAA CTTTTTGCTC TTATAATTTT 1980  
ATTTTATTAA ATTAAATTAC TGCAGTGGAA AAAAAAAAA 2018

(2) INFORMATION FOR SEQ ID NO: 152:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 942 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: LUNGTUT13  
(B) CLONE: 3117184

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152 :

CCTCCATCAG CTCGCCGCGC AGCGGCTGTA TTTGCGGCCT GTGCGAGTAG GCGCTTGGGC 60  
ACTCAGTCTC CCTGGCGGGC GACGGGCAGA AATCTCGAAC CAGTGGAGCG CACTCGTAAC 120  
CTGGATCCCA GAAGGTCGCG AAGGCAGTAC CGTTTCCTCA GCGGCGGACT GCTGCAGTAA 180  
GAATGTCTTT TCCACCTCAT TTGAATCGCC CTCCCATGGG AATCCCAGCA CTCCCACCAG 240  
GGACCCACC CCCGCAGTTT CCAGGATTTT CTCCACCTGT ACCTCCAGGG ACCCCAATGA 300  
TTCCTGTACC AATGAGCATT ATGGCTCCTG CTCCGACTGT CTTAGTACCC ACTGTGTCTA 360  
TGTTTGAAA GCATTTGGGC GCAAGAAAGG ATCATCCAGG CTTAAAGGCT AAAGAAAATG 420  
ATGAAAATTG TGGTCCTACT ACCACTGTTT TTGTTGGCAA CATTTCCGAG AAAGCTTCAG 480  
ACATGCTTAT AAGACAACCT TTAGCTAAAT GTGGTTTGGT TTTGAGCTGG AAGAGAGTAC 540  
AAGGTGCTTC CGGAAAGCTT CAAGCCTTCG GATTCTGTGA GTACAAGGAG CCAGAATCTA 600  
CCCTCCGTGC ACTCAGATTA TTACATGACC TGCAAATTGG AGAGAAAAAG CTA CTACTCGTTA 660



PF-0459 US

AAGTTGATGC AAAGACAAAG GCACAGCTGG ATGAATGGAA AGCAAAGAAG AAAGCTTCTA 720  
ATGGGAATGC AAGGCCAGAA ACTGTCATA ATGACGATGA AGAAGCCTTG GATGAAGAAA 780  
CAAAGAGGAG AGATCAGATG ATTAAAGGGG CTATTGAAGT TTTAATTCGT GAATACTCCA 840  
GTGAGCTAAA TGCCCCCTCA CAGGAATCTG ATTCTCACCC CAGGAAGAAG AAGAAGGAAA 900  
AGAAGGAGGA CATTTTCGGC AGATTTCAGT GGGCCCCACTG AT 942

(2) INFORMATION FOR SEQ ID NO: 153:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2060 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LNODNOT05
- (B) CLONE: 3125156

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153 :

TCCCCCCTC AGCCTCCCC CCCCCACTG GCATATGGTC CTGCCCCCTC TACCAGACCC 60  
ATGGGCCCC AGGCAGCCCC TCTTACCATT CGAGGGCCCT CGTCTGCTGG CCAGTCCACC 120  
CCTAGTCCCC ACCTGGTGCC TTCACCTGCC CCATCTCCAG GGCCTGGTCC GGTACCCCCT 180  
CGCCCCCAG CAGCAGAACC ACCCCCTTGC CTGCGCCGAG GCGCCGCAGC TGCAGACCTG 240  
CTCTCCTCCA GCCCGGAGAG CCAGCATGGC GGCACCTCAGT CTCCTGGGGG TGGGCAGCCC 300  
CTGCTGCAGC CCACCAAGGT GGATGCAGCT GAGGGTCGTC GGCCGCAGGC CCTGCGGCTG 360  
ATTGAGCGGG ACCCCTATGA GCATCCTGAG AGGCTGCGGC AGTTGCAGCA GGAGCTGGAG 420  
GCCTTTCGGG GTCAGCTGGG GGATGTGGGA GCTCTGGACA CTGTCTGGCG AGAGCTGCAA 480  
GATGCGCAGG AACATGATGC CCGAGGCCGT TCCATCGCCA TTGCCCGCTG CTACTIONACTG 540  
AAGAACCGGC ACCAGGATGT CATGCCCTAT GACAGTAACC GTGTGGTGCT GCGCTCAGGC 600  
AAGGATGACT ACATCAATGC CAGCTGCGTG GAGGGGCTCT CCCCATACTG CCCCCCGCTA 660  
GTGGCAACCC AGGCCCCACT GCCTGGCACA GCTGCTGACT TCTGGCTCAT GGTCCATGAG 720  
CAGAAAGTGT CAGTCATTGT CATGCTGGTT TCTGAGGCTG AGATGGAGAA GCAAAAAGTG 780  
GCACGCTACT TCCCCACCGA GAGGGGCCAG CCCATGGTGC ACGGTGCCCT GAGCCTGGCA 840  
TTGAGCAGCG TCCGCAGCAC CGAAACCCAT GTGGAGCGCG TGCTGAGCCT GCAGTTCCGA 900  
GACCAGAGCC TCAAGCGCTC TCTTGTGCAC CTGCACTTCC CCACTTGGCC TGAGTTAGGC 960  
CTGCCCCACA GCCCCAGCAA CTTGCTGCGC TTCATCCAGG AGGTGCACGC ACATTACCTG 1020

PF-0459 US

CATCAGCGGC CGCTGCACAC GCCCATCATT GTGCACTGCA GCTCTGGTGT GGGCCGCACG 1080  
GGAGCCTTTG CACTGCTCTA TGCAGCTGTG CAGGAGGTGG AGGCTGGGAA CGGAATCCCT 1140  
GAGCTGCCTC AGCTGGTGCG GCGCATGCGG CAGCAGAGAA AGCACATGCT GCAGGAGAAG 1200  
CTGCACCTCA GGTTCGTGTA TGAGGCAGTG GTGAGACACG TGGAGCAGGT CCTGCAGCGC 1260  
CATGGTGTGC CTCCTCCATG CAAACCCTTG GCCAGTGCAA GCATCAGCCA GAAGAACCAC 1320  
CTTCCTCAGG ACTCCCAGGA CCTGGTCCTC GGTGGGGATG TGCCCATCAG CTCCATCCAG 1380  
GCCACCATTG CCAAGCTCAG CATTCGGCCT CCTGGGGGGT TGGAGTCCCC GGTGCCAGC 1440  
TTGCCAGGCC CTGCAGAGCC CCCAGGCCCTC CCGCCAGCCA GCCTCCCAGA GTCTACCCCA 1500  
ATCCCATCTT CCTCCCAAAC CCCCTTTCCT CCCCACTACC TGAGGCTCCC CAGCCTAAGG 1560  
AGGAGCCGCC AGTGCCTGAA GCCCCAGCT CGGGGCCCCC CTCCTCCTCC CTGGAATTGC 1620  
TGGCCTCCTT GACCCAGAG GCCTTCTCCC TGGACAGCTC CCTGCGGGGC AAACAGCGGA 1680  
TGAGCAAGCA TAACTTTCTG CAGGCCATA ACGGGCAAGG GCTGCGGGCC ACCCGGCCCT 1740  
CTGACGACCC CCTCAGCCTT CTGGATCCAC TCTGGACACT CAACAAGACC TGAACAGGTT 1800  
TTGCCTACCT GGTCTTACA CTACATCATC ATCATCTCAT GCCCACCTGC CCACACCCAG 1860  
CAGAGCTTCT CAGTGGGCAC AGTCTCTTAC TCCCATTCTT GCTGCCTTTG GCCCTGCCTG 1920  
GCCCAGCCTG CACCCCTGTG GGGTGGAAAT GTACTGCAGG CTCTGGGTCA GGTTCGTCTC 1980  
CTTTATGGGA CCCGACATTT TTCAGCTCTT TGCTATTGAA ATAATAAACC ACCCTGTTCT 2040  
GTGAAAAAAA AAAAAAAG 2060

(2) INFORMATION FOR SEQ ID NO: 154:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2065 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: LUNGTUT12
- (B) CLONE: 3129120

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154 :

CGGGTCCCCG GGTCTGACAG GAGCAGCCTG TGGGCACCGC GGCGGTAGTT GGAGGCGGGA 60  
GAGGGTCCGT AGCCGCGCCG CCCTGCCCCG CCATGGGCCT CCTGTCGGAC CCGGTTCCGCC 120  
GGCGCGCGCT CGCCCGCCTA GTGCTGCGCC TCAACGCGCC GTTGTGCGTG CTGAGCTACG 180  
TGGCGGGCAT CGCCTGGTTC TTGGCGCTGG TTTTCCCGCC GCTGACCCAG CGCACTTACA 240

PF-0459 US

TGTCGGAGAA CGCCATGGGC TCCACCATGG TGGAGGAGCA GTTTGCGGGC GGAGACCGTG 300  
CCCGGGCTTT TGCCCGGGAC TTCGCCGCC ACCGCAAGAA GTCGGGGGCT CTGCCAGTGG 360  
CCTGGCTTGA ACGGACGATG CGGTCAGTAG GGCTGGAGGT CTACACGCAG AGTTTCTCCC 420  
GGAAACTGCC CTTCCCAGAT GAGACCCACG AGCGCTATAT GGTGTCGGGC ACCAACGTGT 480  
ACGGCATCCT GCGGGCCCCG CGTGCTGCCA GCACCGAGTC GCTTGTGCTC ACCGTGCCCT 540  
GTGGCTCTGA CTCTACCAAC AGCCAGGCTG TGGGGCTGCT GCTGGCACTG GCTGCCCCT 600  
TCCGGGGGCA GATTTATTGG GCCAAAGATA TCGTCTTCCT GGTAACAGAA CATGACCTTC 660  
TGGGCACTGA GGCTTGGCTT GAAGCCTACC ACATGTCAA TGTCAGTGGC ATGCAGTCGT 720  
CTCCCCTGCA GGGCCGAGCT GGGGCCATTC AGGCAGCCGT GGCCCTGGAG CTGAGCAGTG 780  
ATGTGGTCAC CAGCCTCGAT GTGGCCGTGG AGGGGCTTAA CGGGCAGCTG CCCAACCTTG 840  
ACCTGCTCAA TCTCTTCCAG ACCTTCTGCC AGAAAGGGGG CCTGTTGTGC ACGCTTCAGG 900  
GCAAGCTGCA GCCCGAGGAC TGGACATCAT TGGATGGACC GCTGCAGGGC CTGCAGACAC 960  
TGCTGCTCAT GGTTCGCGG CAGGCCTCCG GCCGCCCCCA CGGCTCCCAT GGCCTCTTCC 1020  
TGCGCTACCG TGTGGAGGCC CTAACCTTGC GTGGCATCAA TAGCTTCCGC CAGTACAAGT 1080  
ATGACCTGGT GGCAGTGGGC AAGGCTTTGG AGGGCATGTT CCGCAAGCTC AACCACCTCC 1140  
TGGAGCGCCT GCACCAGTCC TTCTTCCTCT ACTTGCTCCC CGGCCTCTCC CGCTTCGTCT 1200  
CCATCGGCCT CTACATGCCC GCTGTCCGCT TCTTGCTCCT GGTCTTGGT CTCAAGGCTC 1260  
TGGAAGTGTG GATGCAGCTG CATGAGGCTG GAATGGGCCT TGAGGAGCCC GGGGGTGCCC 1320  
CTGGCCCCAG TGTACCCCTT CCCCCATCAC AGGGTGTGGG GCTGGCCTCG CTCGTGGCAC 1380  
CTCTGCTGAT CTCACAGGCC ATGGGACTGG CCCTCTATGT CCTGCCAGTG CTGGGCCAAC 1440  
ACGTTGCCAC CCAGCACTTC CCAGTGGCAG AGGCTGAGGC TGTGGTGCTG AACTGCTGG 1500  
CGATTTATGC AGCTGGCCTG GCCCTGCCCC ACAATACCCA CCGGGTGGTA AGCACACAGG 1560  
CCCCAGACAG GGGCTGGATG GCACTGAAGC TGCTAGCCCT GATCTACCTA GCACTGCAGC 1620  
TGGGCTGCAT CGCCCTCACC AACTTCTCAC TGGGCTTCCT GCTGGCCACC ACCATGGTGC 1680  
CCACTGCTGC GCTTGCCAAG CCTCATGGGC CCCGGACCCT CTATGCTGCC CTGCTGGTGC 1740  
TGACCAGCCC GGCAGCCACG CTCCTTGGCA GCCTGTTCCCT GTGGCGGGAG CTGCAGGAGG 1800  
CGCCACTGTC ACTGGCCGAG GGCTGGCAGC TCTTCCTGGC AGCGCTAGCC CAGGGTGTGC 1860  
TGGAGACCA CACCTACGGC GCCCTGCTCT TCCCACTGCT GTCCCTGGGC CTCTACCCCT 1920  
GCTGGCTGCT TTTCTGGAAT GTGCTCTTCT GGAAGTGAGA TCTGCCTGTC CGGGCTGGGA 1980  
CAGAGACTCC CCAAGGACCC CATTCTGCCT CCTTCTGGGG AAATAAATGA GTGTCTGTTT 2040  
CAGCAGCTAT TTGATGCTTG TCACA 2065

What is claimed is:

1. A substantially purified human signal peptide-containing protein (SIGP) comprising a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, and SEQ ID NO:77.

2. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding an SIGP of claim 1.

3. An isolated and purified polynucleotide encoding the SIGP of claim 1.

4. A microarray containing at least a fragment of at least one of the polynucleotides encoding an SIGP of claim 1.

5. An isolated and purified polynucleotide variant having at least 90% polynucleotide identity to the polynucleotide of claim 3.

6. A composition comprising the polynucleotide of claim 3.

7. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.

8. An isolated and purified polynucleotide which is complementary to the polynucleotide of claim 3.

9. An isolated and purified polynucleotide having a nucleic acid sequence

selected from the group consisting of SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, and SEQ ID NO:154.

10. An isolated and purified polynucleotide variant having at least 90% polynucleotide identity to the polynucleotide of claim 9.

11. An isolated and purified polynucleotide which is complementary to the polynucleotide sequence of claim 9.

12. An expression vector containing at least a fragment of the polynucleotide of claim 3.

13. A host cell containing the expression vector of claim 12.

14. A method for producing a polypeptide encoding a human signal peptide-containing protein, the method comprising the steps of:

(a) culturing the host cell of claim 13 under conditions suitable for the expression of the polypeptide; and

(b) recovering the polypeptide from the host cell culture.

15. A pharmaceutical composition comprising the SIGP of claim 1 in conjunction with a suitable pharmaceutical carrier.

16. A purified antibody which specifically binds to the SIGP of claim 1.

17. A purified agonist of the SIGP of claim 1.

18. A purified antagonist of the SIGP of claim 1.

5 19. A method for treating or preventing a cancer, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 15.

20. A method for treating or preventing a cancer, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.

10 21. A method for treating or preventing an immune response, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.

15 22. A method for detecting a polynucleotide encoding a human signal peptide-containing protein in a biological sample containing nucleic acids, the method comprising the steps of:

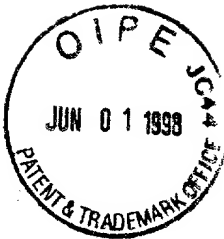
(a) hybridizing the polynucleotide of claim 8 to at least one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding SIGP in the biological sample.

20 23. The method of claim 22 wherein the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.

## ABSTRACT OF THE DISCLOSURE

The invention provides a human signal peptide-containing proteins (SIGP) and  
5 polynucleotides which identify and encode SIGP. The invention also provides expression  
vectors, host cells, antibodies, agonists, and antagonists. The invention also provides  
methods for treating or preventing disorders associated with expression of SIGP.



Docket No.: PF-0459 US

**DECLARATION AND POWER OF ATTORNEY FOR  
UNITED STATES PATENT APPLICATION**

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,  
and

I believe that I am the original, first and sole inventor (if only one name is listed below)  
or an original, first and joint inventor (if more than one name is listed below) of the subject  
matter which is claimed and for which a United States patent is sought on the invention entitled

**HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS**

the specification of which:

   / is attached hereto.

  X   / was filed on December 31, 1997, as application Serial No. 09/002,485 and if this box  
contains an X    /, was amended on \_\_\_\_\_.

   / was filed as Patent Cooperation Treaty international application No. \_\_\_\_\_ on  
\_\_\_\_\_, 19\_\_\_\_, if this box contains an X    /, was amended on under Patent Cooperation  
Treaty Article 19 on \_\_\_\_\_ 19\_\_\_\_, and if this box contains an X    /, was amended on \_\_\_\_\_.

I hereby state that I have reviewed and understand the contents of the above-identified  
specification, including the claims, as amended by any amendment referred to above.

I acknowledge my duty to disclose information which is material to the examination of  
this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim the benefit under Title 35, United States Code, §119 or §365(a)-(b) of any  
foreign application(s) for patent or inventor's certificate indicated below and of any Patent  
Cooperation Treaty international applications(s) designating at least one country other than the  
United States indicated below and have also identified below any foreign application(s) for patent  
or inventor's certificate and Patent Cooperation Treaty international application(s) designating at  
least one country other than the United States for the same subject matter and having a filing date  
before that of the application for said subject matter the priority of which is claimed:



Country	Number	Filing Date	Priority Claimed
			// Yes    // No
			// Yes    // No

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of Title 35, United States Code §112, I acknowledge my duty to disclose material information as defined in Title 37 Code of Federal Regulations, §1.56(a) which occurred between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)

I hereby appoint the following:

<b>LUCY J. BILLINGS</b>	<b>Registration No. 36,749</b>
<b>MICHAEL C. CERRONE</b>	<b>Registration No. 39,132</b>
<b>SHEELA MOHAN-PETERSON</b>	<b>Registration No. 41,201</b>
<b>COLETTE C. MUENZEN</b>	<b>Registration No. 39,784</b>
<b>KAREN J. ZELLER</b>	<b>Registration No. 37,071</b>

respectively and individually, as my attorneys and/or agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith. Please address all communications to:

**MICHAEL C. CERRONE, Ph.D.**  
**INCYTE PHARMACEUTICALS, INC.**  
**3174 PORTER DRIVE, PALO ALTO, CA 94304**

**TEL: 650-855-0555      FAX: 650-845-4166**

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are

punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

**Sole Inventor or**

**First Joint Inventor:**

Full name:

**PREETI LAL**

Signature:

Preeti Lal

Date:

15<sup>th</sup> May 1998

Citizenship:

India

Residence:

Santa Clara, California

P.O. Address:

2382 Lass Drive  
Santa Clara, CA 95054

**Scnd Joint Inventor:**

Full name:

**JENNIFER L. HILLMAN**

Signature:

Jennifer L. Hillman

Date:

May 13, 1998

Citizenship:

United States of America

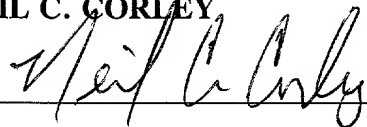
Residence:

Mountain View, California

P.O. Address:

230 Monroe Drive, #12  
Mountain View, CA 94040

**Third Joint Inventor:**

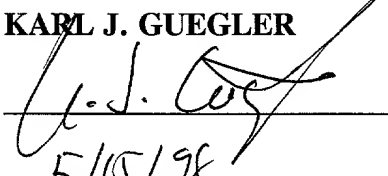
Full name: **NEIL C. CORLEY**  
Signature:   
Date: 5-19-98

Citizenship: United States of America

Residence: Mountain View, California

P.O. Address: 1240 Dale Avenue, #30  
Mountain View, CA 94040

**Fourth Joint Inventor:**


Full name: **KARL J. GUEGLER**  
Signature:   
Date: 5/15/98

Citizenship: Switzerland

Residence: Menlo Park, California

P.O. Address: 1048 Oakland Avenue  
Menlo Park, CA 94025

**Fifth Joint Inventor:**

Full Name: <sup>BAUGHN MB</sup>  
**MARIAH R. BAUGH**  
Signature:   
Date: 5/19/98

Citizenship: United States of America

Residence: San Jose, California

P.,O. Address: 1537 LaRossa Circle  
San Jose, CA 95125

**Sixth Joint Inventor:**

Full Name: **SUSAN K. SATHER**  
Signature: *Susan K. Sather*  
Date: 5/18/98  
Citizenship: United States of America  
Residence: Palo Alto, California  
P.O. Address: 2721 Midtown Court, Apt. #203  
Palo Alto, CA 94303

**Seventh Joint Inventor:**

Full name: **PURVI SHAH**  
Signature: *Purvi Shah*  
Date: 5/15/98  
Citizenship: India  
Residence: Sunnyvale, California  
P.O. Address: 1608 Queen Charlotte Drive, #5  
Sunnyvale, CA 94087